Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Strength & Weakness of Lee-Carter Models

Li Yang

MSc 08/2019-014

School of Business and Economics

Master thesis *in*
Economics & Financial Research track Econometrics

Strength & Weakness of Lee-Carter Models

Abstract

This paper investigates the identifiability and estimation method of the Lee-Carter model and its several extensions. In particular, it focuses on the estimation method of the plug-in Lee-Carter models. The main finding of this paper is that after adjusting the proposed objective functions for the plug-in Lee-Carter models, the goodness of fit is improved.

Student: Li Yang
Student ID: i6091224
Supervisor: Dr. Eric Beutner

Date submitted: 26 August 2019

CONTENTS

1 Introduction

Historical data show that mortality rates have been decreasing during last decades. A consequence of this development is that life expectancy, that how long the individual is expected to live, has been continuously going beyond previously-held limit, which is known as longevity risk. For instance, longevity risk is a key risk affecting the costs of private and public health insurance, as well as public programs of old-age support and pensions. Unlike diversifiable mortality risks, longevity risk cannot be diversified by holding a large number of polices, because it affects the entire portfolios. To cushion against this risk, risk capital is often required. One important ingredient needed to determine the amount of this capital is future mortality rates, which can be forecast by mortality models. Among those, the Lee-Carter model is one commonly used mortality model. This model was introduced by Lee & Carter (1992). The Lee-Carter model has the following form:

$$ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \ x \in [0, ..., X], \ t \in [1, .., T]. \tag{1}$$

$m_{x,t}$ is the central death rate for age x. To ensure the identifiability of this model, that the model parameters can be estimated uniquely, Lee & Carter (1992) proposed the following constraints: $\sum_{x=0}^{X} \beta_x = 1, \ \sum_{t=1}^{T} \kappa_t = 0,$ and implicitly assumed that $E(\epsilon_{x,t}) = 0, \ Var(\epsilon_{x,t}) = \sigma_\epsilon^2$. Under these constraints, α_x maps the main age profile of mortality, κ_t represents the period effects and β_x measures the interactions with age.

The Lee-Carter model has a number of advantages, among which are its simplicity and that the parameters are parsimonious and easily interpretable. Additionally, the model fits US data for the period 1933-1987 and the G7 countries for the period 1950-1995 very well (Tuljapurkar et al., 2000). However, there are shortcomings of this model. The Lee-Carter model assumes that the errors are homoskedastic, which is quite unrealistic because the variance of the logarithm of the observed force of mortality is much greater at older ages than at younger ages, pointed out by the paper Brouhns et al. (2002). Furthermore, the Lee-Carter model assumes that the sensitivity of log mortality rates at each age (β_t) remains constant, while the paper by Booth, Maindonald & Smith (2002) shows that age-time interaction is highly likely in Australian data. In addition, the paper by Renshaw & Haberman (2006) claims that incorporation of cohort (age-time interaction) effects improves the performance of the Lee-Carter model. Recently, the paper by Leng & Peng (2016) points out that the two-step procedure may lead to inconsistent estimator. Moreover, the paper by Beutner, Reese & Urbain (2017) states that the identification constraint $\sum_{t=1}^{T} \kappa_t = 0$ becomes a constraint on the possible realization of the stochastic process $\{\kappa_t\}$, therefore, it proposes a so-called plug-in Lee-Carter model where the time-varying variable κ_t is replaced by time series models from the beginning. These last two papers will be the focus of this paper.

This paper examines the strength and the weakness of the Lee-Carter model and its several extensions. It is organized as follows. In section two, it examines the identifiability and the estimation method of the Lee-Carter model and gives real data analysis. Two extensions of this model will be introduced in section three. In particular, the identifiability and the estimation methods of these models will be investigated and real data analysis will be presented. In section four, the pitfalls of employing the Lee-Carter estimation procedure will be shown and a simulation study will be given to support the theoretical findings. Two more extensions of the Lee-Carter model, the plug-in age-period Lee-Carter model and the plug-in age-period-cohort Lee-Carter model will be introduced in section five. Its estimation and forecasting methods will be the focus of this section and real data analysis will be given. Finally, conclusions will be drawn in section six.

2 The Lee-Carter model

2.1 Identifiability of the Lee-Carter model

In order to estimate the parameters, a model is required to be identifiable. The definition of identifiability is shown below.

Definition of identifiability Let $x, y \in \{0, \cdots, X\}$, $s, t \in \{1, \cdots, X\}$, $\theta, \tilde{\theta} \in \Theta$ with the parameter space Θ being a subset of some finite dimensional space. Furthermore,

$$E_\theta\big(ln(m_{x,t})\big) = f_\theta(x,t), Cov_\theta\big(ln(m_{x,s}), ln(m_{y,t})\big) = g_\theta(x,y,s,t).$$
$$E_{\tilde{\theta}}\big(ln(m_{x,t})\big) = f_{\tilde{\theta}}(x,t), Cov_{\tilde{\theta}}\big(ln(m_{x,s}), ln(m_{y,t})\big) = g_{\tilde{\theta}}(x,y,s,t).$$

If $f_\theta = f_{\tilde{\theta}}$ implies $\theta = \tilde{\theta}$, then the expected values are identifiable;
If $f_\theta = f_{\tilde{\theta}},\ g_\theta = g_{\tilde{\theta}}$ together imply $\theta = \tilde{\theta}$, then the expected values and the covariance structure are identifiable. (Beutner, Reese & Urbain, 2017)

As mentioned earlier, the constraints which can ensure the identifiabiltiy of the Lee-Carter model, are as follows:

$$\sum_{x=0}^{X} \beta_x = 1,\ \sum_{t=1}^{T} \kappa_t = 0. \tag{2}$$

.

Proof Let $\theta = \{\alpha, \beta, \kappa\},\ \tilde{\theta} = \{\tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}\},\ \theta,\ \tilde{\theta} \in \Theta$. Assume $E\big(ln_\theta(m_{x,t})\big) = E\big(ln_{\tilde{\theta}}(m_{x,t})\big)$, then

$$\alpha_x + \beta_x\kappa_t = \tilde{\alpha}_x + \tilde{\beta}_x\tilde{\kappa}_t,$$
$$\text{it implies } \sum_{t=1}^{T}(\alpha_x + \beta_x\kappa_t) = \sum_{t=1}^{T}(\tilde{\alpha}_x + \tilde{\beta}_x\tilde{\kappa}_t),$$
$$T\alpha_x + \beta_x\sum_{t=1}^{T}\kappa_t = T\tilde{\alpha}_x + \tilde{\beta}_x\sum_{t=1}^{T}\tilde{\kappa}_t,$$
$$\alpha_x \overset{(2)}{=} \tilde{\alpha}_x,$$
therefore,
$$\beta_x\kappa_t = \tilde{\beta}_x\tilde{\kappa}_t,$$
$$\text{it implies } \sum_{x=0}^{X}(\beta_x\kappa_t) = \sum_{x=0}^{X}(\tilde{\beta}_x\tilde{\kappa}_t),$$
$$\left(\sum_{x=0}^{X}\beta_x\right)\kappa_t = \left(\sum_{x=0}^{X}\tilde{\beta}_x\right)\tilde{\kappa}_t,$$
$$\kappa_t \overset{(2)}{=} \tilde{\kappa}_t.$$
If there exits a κ_t such that $\kappa_t \neq 0$, then
$$\beta_x = \tilde{\beta}_x,$$
$$\theta = \tilde{\theta}.$$

As a result, the expected values of the Lee-Carter model are identifiable.

2.2 Estimation of the Lee-Carter Model

Step 1:
The parameters are estimated as follows.

$$
\begin{aligned}
ln(m_{x,t}) &= \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \\
\sum_{t=1}^{T} ln(m_{x,t}) &= \sum_{t=1}^{T} (\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}) \\
\sum_{t=1}^{T} ln(m_{x,t}) &= \sum_{t=1}^{T} \alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t + \sum_{t=1}^{T} \epsilon_{x,t} \\
&\approx T\alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t \\
&\stackrel{(2)}{=} T\alpha_x
\end{aligned}
$$

$$\hat{\alpha}_x = \frac{1}{T} \sum_{t=1}^{T} ln(m_{x,t}) \tag{3}$$

let

$$A = \begin{pmatrix} ln(m_{0,1}) - \hat{\alpha}_0 & ln(m_{0,2}) - \hat{\alpha}_0 & \cdots & ln(m_{0,T}) - \hat{\alpha}_0 \\ ln(m_{1,1}) - \hat{\alpha}_1 & ln(m_{1,2}) - \hat{\alpha}_1 & \cdots & ln(m_{1,T}) - \hat{\alpha}_1 \\ \vdots & \vdots & \ddots & \vdots \\ ln(m_{X,1}) - \hat{\alpha}_X & ln(m_{X,2}) - \hat{\alpha}_X & \cdots & ln(m_{X,T}) - \hat{\alpha}_X \end{pmatrix}, \tag{4}$$

apply Singular Value Decomposition (SVD) to the matrix A, which is to factorize A into the product of three matrices $A = U\Sigma V^T$, where the columns of U and V are orthogonormal and the matrix Σ is diagonal with positive real entries. Here, Σ is an $(X+1) \times T$ diagonal matrix with singular values (SV) on the diagonal. In case $rank(A) = r \leq T \leq X+1$, the matrix Σ can be showed in following:

$$\Sigma = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \lambda_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_{r+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \lambda_T \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{5}$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0,\ \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_T = 0$. Particularly, A can be written in the following form:

$$A = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \cdots + \lambda_r u_r v_r^T. \tag{6}$$

The $u_i's$ and $v_i's$ are called the left and the right singular-vectors of A. To approximate β and κ,

$\hat{\beta} = \frac{u_1}{\sum_{i=1}^{X+1} u_{1,i}}$, namely the first left vector normalized,

$\hat{\kappa} = v_1 \times \big(\sum_{i=1}^{X+1} u_{1,i} \big) \times \lambda_1$, namely the first right vector modified, λ_1 is the leading value of the SV.

Note that if $\sum_{i=1}^{X+1} u_{1,i} = 0$, $\hat{\beta}$ cannot be normalized. Furthermore, the components of v_1 will always be 0, which will be shown later, hence, $\sum_t^T \hat{\kappa}_t = 0$ always holds.

Step 2:
Take $\hat{\alpha}$, $\hat{\beta}$ from the first step, re-estimate $\kappa_t' s$ such that they predict the correct total number of deaths each year, namely, $\sum_x D_{x,t} = \sum_x e_{x,t} exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\hat{\kappa}}_t)$, $e_{x,t}$ is the number of person years from which $D_{x,t}$ occurred.

Note that, under the constraint $\sum_t \kappa_t = 0$ and the assumption that the error terms are independent and identically distributed (i.i.d) with distribution $N(0,1)$, the estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\kappa}$ are also maximum likelihood estimates, which will be shown in later section. Furthermore, ARIMA(p,d,q) models are applied to $\{\hat{\kappa}_t\}$ and $\{\hat{\hat{\kappa}}_t\}$ to forecast $\kappa_{T+1}, \kappa_{T+2}$, etc. In addition, p is the number of autoregressive terms, d the number of differences needed for stationarity and q the number of lagged forecast errors in the prediction equation. In practice, ARIMA(p,1,q) is preferred.

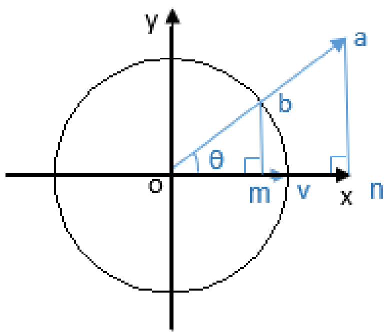
2.2.1 SVD & Least squares

Let A,U,V and Σ be the elements such that $A = U \Sigma V^T$. $A = (\vec{a_1}, ..., \vec{a_m})$ and Σ is a diagonal matrix with λ_i as its ith diagonal entry. Furthermore, define

$$\bar{v} = arg\ max_{v \in R^n,\ ||v||=1} ||AV||. \tag{7}$$

Let $\vec{v}$ be an arbitrary unit vector, $\vec{a}$ be an arbitrary vector, and θ is the acute angle formed by these two vectors, shown in Figure 1. Furthermore, $\vec{n}$ is the projection of $\vec{a}$ on vector $\vec{v}$.

Figure 1: Projection of $\vec{a}$ on $\vec{v}$.



Then

$$\vec{a} \cdot \vec{v} = \sum_{j=1}^{n} a_j v_j = cos(\theta) \times ||\vec{a}|| \times ||\vec{v}|| = ||\vec{a}|| cos(\theta) = ||\vec{n}||, \tag{8}$$

therefore, $\sum_i^n \sum_j^n a_{ij}v_j$ is the sum of the lengths of the projections of $\vec{a_i}$ on $\vec{v}$. As a result,

$$||A\vec{v}|| = \sqrt{\sum_{i=1}^{m}(\sum_{j}^{n} a_{ij}v_j)^2} = \sqrt{\sum_{i=1}^{m} ||\vec{a}_i||^2 cos(\theta_i)^2} \tag{9}$$

$$= \sqrt{\sum_{i=1}^{m} ||\vec{a}_i||^2 - \sum_{i=1}^{m} ||\vec{a}_i||^2 sin(\theta_i)^2}. \tag{10}$$

Maximizing $||A\vec{v}||$ is the same as maximizing $||A\vec{v}||^2$. Since $||\vec{a}_i||^2$ is constant, maximizing $\sum_{i=1}^m ||\vec{a}_i||^2 cos(\theta_i)^2$ is the same as minimizing $\sum_{i=1}^m ||\vec{a}_i||^2 sin(\theta_i)^2$, which is the sum of the squared distance of $\vec{a}_i$ to $\vec{v}$.
As $A = U\Sigma V^T$, then

$$A^T A = V\Sigma^2 V^T. \tag{11}$$

Now let $\vec{v_1}, ..., \vec{v_m}$ be the normalized eigen vectors of $A^T A$. By definition of $\bar{v}$,

$$||A\vec{v_i}|| \leq ||A\bar{v}||,$$

moreover, for any unit vector $\vec{v}$,

$$\begin{aligned}
||A\vec{v}|| &= \sqrt{\vec{v}^T A^T A\vec{v}} \\
&= \sqrt{(\sum_{i}^{m} c_i\vec{v}_i^T)A^T A(\sum_{i=1}^{m} c_i\vec{v}_i)}, \quad where\ \vec{v} = \sum_{i=1}^{m} c_i\vec{v}_i \\
&= \sqrt{(\sum_{i}^{m} c_i\vec{v}_i^T)(\sum_{i=1}^{m} c_i A^T A\vec{v}_i)} \\
&= \sqrt{(\sum_{i}^{m} c_i\vec{v}_i^T)(\sum_{i=1}^{m} c_i \lambda_i^2\vec{v}_i)} \\
&= \sqrt{\sum_{i}^{m} c_i^2\lambda_i^2}, \quad as\ \vec{v}_i^T\vec{v}_j = 0, if\ i \neq j \\
&\leq \lambda_1 = \sqrt{1 \times \lambda_1^2 + 0 \times \lambda_2^2 + ... + 0 \times \lambda_n^2}.
\end{aligned}$$

Note that the upper bound is obtained, because

$$\begin{aligned}
||A\vec{v}_1|| &= \sqrt{\vec{v}_1^T A^T A\vec{v}_1} \\
&= \sqrt{\vec{v}_1^T \lambda_1^2\vec{v}_1} \\
&= \lambda_1.
\end{aligned}$$

In short, the sum of squared distances of $\vec{a}_i$ (for all i) to an arbitrary unit vector $\vec{v}$, $\sum_{i=1}^m ||\vec{a}_i||^2 sin(\theta_i)^2$, is minimized when $\vec{v}_1$ is the unit vector. Furthermore, the sum of projection of $\vec{a}_i$ (for all i), $\sum_{i=1}^m ||\vec{a}_i||^2 cos(\theta_i)^2$, is maximized on $\vec{v}_1$ with a value of λ_1^2. (Beutner , 2019)

2.2.2 SVD & Lee-Carter estimation

It is mentioned earlier that $\sum_t^T \hat{\kappa}_t = 0$ always holds and it will be shown in this section. Known from equation (3) and (4), we know that the row sum of A is 0, namely, $\sum_{i=1}^n a_{li} = 0, \ \forall l$. Let (i,j)th element of $A^T A$ is given by $\sum_{l=1}^m a_{li} a_{lj}$, then

$$\begin{aligned}
\sum_{i=1}^n \sum_{l=1}^m a_{li} a_{lj} &= \sum_{l=1}^m \sum_{i=1}^n a_{li} a_{lj} \\
&= \sum_{l=1}^m a_{lj} \sum_{i=1}^n a_{li} \\
&= 0,
\end{aligned}$$

which implies that the column sum of $A^T A$ is 0.
Now let B be an $n \times n$ matrix with column sums equal to 0, namely, $\sum_{i=1}^n b_{ij} = 0$. For an arbitrary vector $\vec{v}$,

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n b_{ij} v_j &= \sum_{j=1}^n \sum_{i=1}^n b_{ij} v_j \\
&= \sum_{j=1}^n v_j \sum_{i=1}^n b_{ij} \\
&= 0,
\end{aligned}$$

that is the components of $\vec{v}$ under B sum to 0.
For an eigenvector $\vec{v}$ of B to the eigenvalue $\lambda \neq 0$,

$$\begin{aligned}
B\vec{v} &= \lambda \vec{v} \\
\sum_{j=1}^n \sum_{i=1}^n b_{ij} v_j &= \sum_{j=1}^n \lambda v_j.
\end{aligned}$$

Note that $\hat{\kappa} = v_1 \times \left(\sum_{i=1}^{X+1} u_{1,i} \right) \times \lambda_1$, therefore, $\sum_t^T \hat{\kappa}_t = 0$. (Beutner , 2019)

2.2.3 Least squares & Lee-Carter estimation

If assume the errors are independent and identically distributed (i.i.d) with distribution $N(0, 1)$, the likelihood of the observations is:

$$L = \prod_{x=0}^X \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} exp(- \frac{\big((ln(m_{x,t}) - (\alpha_x + \beta_x \kappa_t) \big)^2}{2}),$$

then the loglikelihood (constant dropped) is:

$$l = \sum_{x=0}^X \sum_{t=1}^T - \frac{1}{2} \big(ln(m_{x,t}) - (\alpha_x + \beta_x \kappa_t) \big)^2.$$

Hence, under the constraint $\sum_{t=1}^{T} \kappa_t = 0$,

$$\begin{aligned}\frac{\partial l}{\partial \alpha_{\bar{x}}} &= \sum_{t=1}^{T} (ln(m_{\bar{x},t}) - \alpha_{\bar{x}} - \beta_{\bar{x}} \kappa_t) \\ &= \sum_{t=1}^{T} ln(m_{\bar{x},t}) - \sum_{t=1}^{T} \alpha_{\bar{x}} - \beta_{\bar{x}} \sum_{t=1}^{T} \kappa_t \\ &= \sum_{t=1}^{T} ln(m_{\bar{x},t}) - T\alpha_{\bar{x}}.\end{aligned}$$

As a result, $\hat{\alpha}_{\bar{x}} = \frac{1}{T} \sum_{t=1}^{T} ln(m_{\bar{x},t})$. Furthermore, maximize the likelihood L is similar to minimize $\sum_{x=0}^{X} \sum_{t=1}^{T} \frac{1}{2} \big(ln(m_{x,t}) - (\alpha_x + \beta_x \kappa_t)\big)^2$, therefore, the Lee-Carter estimation is justified by maximum likelihood estimation. However, the i.i.d. (independent and identical distributed) assumption is not appropriate, for instance, $ln(m_{40,2018})$ is related to $ln(m_{41,2019})$ in such a way that people who aged 40 and survived in 2018 turn 41 in 2019. Moreover, $\hat{\kappa}$ from the first step of the Lee-Carter estimation is the maximum likelihood estimator, while $\hat{\hat{\kappa}}$ resulting from the second step is not. (Beutner , 2019)

2.3 Application of Lee-Carter model

2.3.1 Mortality forecasting

In this section, the Lee-Carter model will be applied to the Dutch life table Total 1x1, 1850-2016 inclusive, with age classified by individual year, 0-100 inclusive. The data are retrieved from the website https://www.mortality.org. The first five observations are displayed in Table 1.

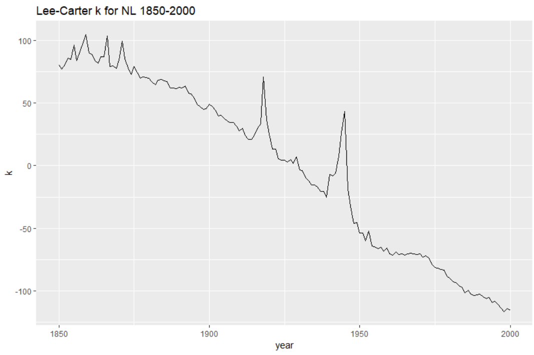
Table 1: First five observations of the Dutch life table Total 1x1.

Year	age	mx	expo	deaths
1850	0	0.22385	87284.26	19539
1850	1	0.06581	72121.34	4746
1850	2	0.03275	63016.54	2064
1850	3	0.01967	62285.37	1225
1850	4	0.01469	66827.91	982

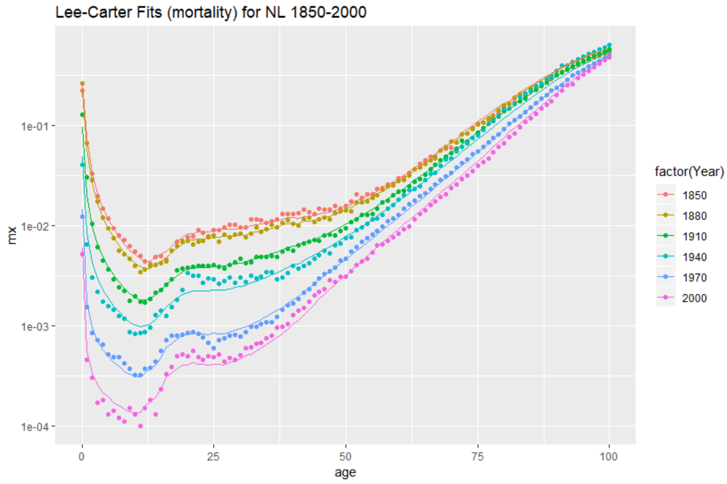
[1] mx: mortality rates, expo: exposure to risk.

Divide it into two subsets such that the training set contains observations before the year 2001 and the test set contains the rest observations. Then estimate α, β and κ by going through the Lee-Carter estimation step 1. The plot of the estimated κ and that of fitted values of mortalities for several years are shown in Figure 2. The plot shows that the Lee-Carter model fits the data very well. To point out that in the plot $E(m_{x,t})$ is set equal to $e^{\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t}$, while $E(m_{x,t}) = e^{\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t + \frac{1}{2}\sigma^2}$. Moreover, it is possible to avoid localized age induced anomalies by smoothing the β_x, for instance, use the method of least squares to fit either cubic B-splines with r knots or natural cubic splines to the estimated β_x (Renshaw & Haberman, 2003). However, this method results in increased number of estimators.

Figure 2: Plot of estimation results.



(a)



(b)

Note that the estimation of the Lee-Carter approach is based on the first set of singular vectors, therefore the variance the first SVD components account for could differ significantly, which could affect the goodness of fit. Let data A be the Dutch life table Total 1x1, 1950-1998 inclusive, with age classified by individual year, 22-60 inclusive, data B be the Dutch life table Male 1x1 with the same specifications. The first five singular values $\{s_i : i = 1, 2, 3, 4, 5\}$ (together with their proportions of the whole, as measured by the ratios $\frac{s_i^2}{\sum_{\text{all i}} s_i^2}$), for these two data sets, are as stated in Table 2. Table 2 shows that the first singular value accounts for 90.6% and 82.1% of the variance, respectively.

Table 2: First five singular values s_i, with $(\frac{s_i^2}{\sum_{\text{all i}} s_i^2})$ as a percentage.

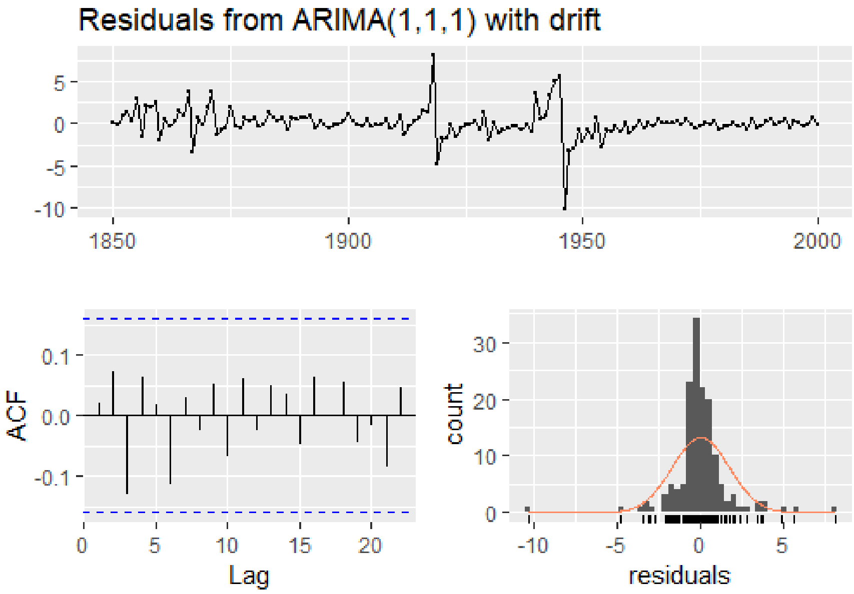
data	components				
	1st	2nd	3rd	4th	5th
A. NL	6.383(90.6%)	1.572(5.5%)	0.835(1.5%)	0.748(1.2%)	0.711(1.1%)
B. NL Male	5.939(82.1%)	2.223(11.5%)	1.07(2.7%)	0.948(2.1%)	0.847(1.7%)

Use the command of **auto.arima** in the R package **forecast** to pick up the best ARIMA(p,d,q) model for $\{\hat{\kappa}_t\}$ by BIC. ARIMA(1,1,1) is picked to be the best model. It is as follows.

$$\begin{array}{cccccccc} \kappa_t & = & \text{-1.4387} & + & 0.7665\ \kappa_{t-1} & + & \text{-0.9434}e_{t-1} & + & e_t \\ & & (0.1934) & & (0.0834) & & (0.0483) & & (8.4994) \end{array}$$

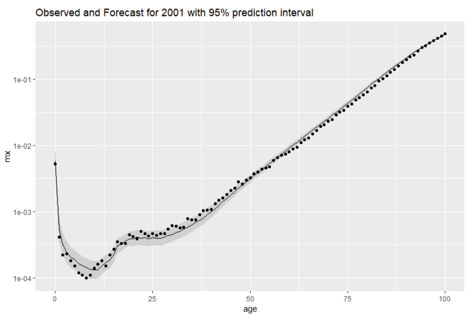
In the parentheses are automatically calculated standard errors, however, they are wrong, because in this step the variability in obtaining $\hat{\kappa}_t$'s is ignored, pointed out by the paper Leng & Peng (2016). The Ljung-Box test with lag=5, df=3 gives a p-value of 0.1023, which implies that the hypothesis that the residuals are not autocorrelated cannot be rejected. The residuals from ARIMA(1,1,1) with drift are shown in Figure 3.

Figure 3: Residuals from ARIMA(1,1,1) with drift.

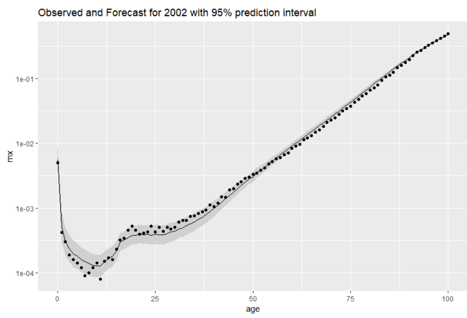


Moreover, the majority of the forecast log mortality fall inside the 95% prediction intervals for the period from 2001 to 2016. Note that these prediction intervals are individual prediction intervals. Figure 4 shows the forecasts for the year 2001, 2002, 2003 and 2016.

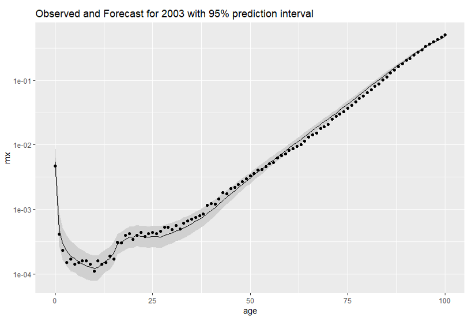
Figure 4: Plot of forecasting results.



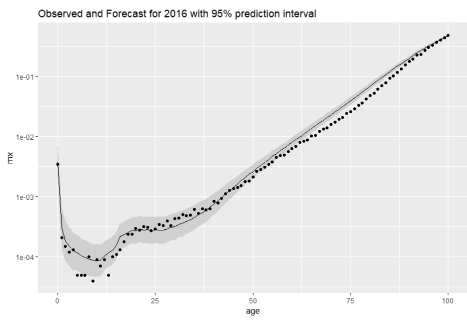
(a) Year 2001 NL.



(b) Year 2002 NL.



(c) Year 2003 NL.



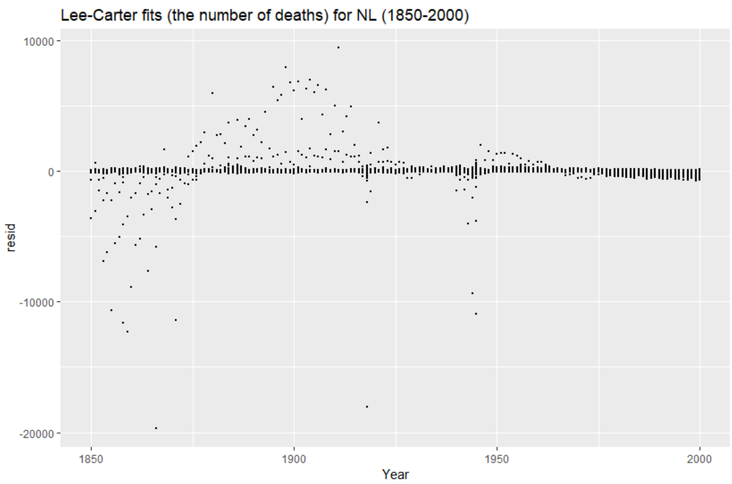
(d) Year 2016 NL.

3 The Lee-Carter model with Poisson errors and its extension

3.1 The PB modeling

Take $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\kappa}$ from the last section and plot in Figure 5 the difference between the actual total deaths and the expected total deaths for each year, then there are curved patterns in such plot, which implies a lack of fit. This is because the Lee-Carter model shown in (1) is fitted to the mortality rates rather than the actual total deaths. The paper Lee & Carter (1992) proposed to re-estimate κ based on the estimates of α and β, which is illustrated in estimation of the Lee-Carter model step 2.

Figure 5: Residual plots against year.



Instead, the paper by Brouhns et al. (2002) suggests modelling D_{xt} (the actual deaths) as independent Poisson response variable, with systematic component:

$$\begin{gathered} E(D_{x,t}) = E_{x,t}exp(\eta_{x,t}), \text{ with } x = 0, \ldots, X, \; t = 1, \ldots, T, \\ \text{where } \eta_{x,t} = \alpha_x + \beta_x\kappa_t, \; E_{x,t} \text{ is the exposure to risk in year t.} \end{gathered} \tag{12}$$

This approach is known as Bilinear predictor, also as PB modelling. Following GLM tradition Brouhns et al. (2002) propose to estimate the parameters by maximizing the log-likelihood, i.e.

$$\sum_{x=0}^{X}\sum_{t=1}^{T} log\left(\frac{\big(E_{x,t}exp(\alpha_x + \beta_x\kappa_t)\big)^{D_{x,t}}}{D_{x,t}!} \times exp(-E_{x,t}exp\big(\alpha_x + \beta_x\kappa_t)\big)\right),$$

which is equal to (up to a constant)

$$\sum_{x=0}^{X}\sum_{t=1}^{T}(\alpha_x + \beta_x\kappa_t)D_{x,t} - E_{x,t}exp\big(\alpha_x + \beta_x\kappa_t\big).$$

Take the first derivative w.r.t. $\alpha_{\tilde{x}}$, $\tilde{x} \in \{0, \ldots, X\}$, and equate it to zero

$$\sum_{t=1}^{T}[D_{\tilde{x},t} - E_{x,t}exp\big(\alpha_{\tilde{x}} + \beta_{\tilde{x}}\kappa_t\big)] = 0,$$

which implies that total actual deaths is equal to total expected deaths for each year. This is required under the Lee-Carter approach and achieved only through re-estimation of κ. In more detail, the parameter set $\Theta := \{(\alpha, \beta, \kappa) \in R^{X+1} \times R^{X+1} \times R^T\}$ are estimated by employing an iterative algorithm which minimizes value of the total deviance of this model, which is given by:

$$\begin{aligned} D(d_{x,t}, \hat{d}_{x,t}) &= -2log\frac{\text{maximum likelihood for model}}{\text{maximum likelihood for saturated model}} \\ &= -2[L(\hat{d}_{x,t}; d_{x,t}) - L(d_{x,t}; d_{x,t})] \\ &= \sum_{\text{all } x,t} 2\{d_{x,t}log\frac{d_{x,t}}{\hat{d}_{x,t}} - d_{x,t} + \hat{d}_{x,t}\}, \end{aligned} \tag{13}$$

where

$$\hat{d}_{x,t} = E_{x,t}exp(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t);$$

Moreover, Butt et al. (2015) suggest that the Newton-Raphson minimization method can be applied to the deviance function. This results in:

$$updated(\hat{\theta}) = \hat{\theta} - \frac{\frac{\partial D}{\partial \theta}}{\frac{\partial^2 D}{\partial \theta^2}}. \tag{14}$$

In addition,

$$\begin{aligned} \frac{\partial D}{\partial \theta} = \sum_{\text{all } x,t} \frac{\partial dev}{\partial \theta_{x,t}} &= \sum 2\omega_{x,t}\{-d_{x,t}\frac{\hat{d}'_{x,t}}{\hat{d}_{x,t}} + \hat{d}'_{x,t}\} \\ &= \sum 2\omega_{x,t}\frac{\hat{d}'_{x,t}}{\hat{d}_{x,t}}(\hat{d}_{x,t} - d_{x,t}) = \sum_{\text{all } x,t} 2\omega_{x,t}a(\hat{d}_{x,t} - d_{x,t}), \end{aligned} \tag{15}$$

where,

$$\hat{d}'_{x,t} = \frac{\partial \hat{d}_{x,t}}{\partial \theta_{x,t}} \rightarrow \begin{cases} \frac{\partial \hat{d}_{x,t}}{\partial \alpha_x} = \hat{d}_{x,t} \\ \frac{\partial \hat{d}_{x,t}}{\partial \beta_x} = \kappa_t \hat{d}_{x,t} \\ \frac{\partial \hat{d}_{x,t}}{\partial \kappa_t} = \beta_x \hat{d}_{x,t} \end{cases} \text{, such that } a = \begin{cases} 1 \\ \kappa_t \\ \beta_x \end{cases} \text{, and } \omega_{x,t} = \begin{cases} 1, \text{ if } e_{x,t} > 0 \\ 0, \text{ otherwise} \end{cases}.$$

Hence,

$$\frac{\partial^2 D}{\partial \theta^2} = \sum_{\text{all } x,t} 2\omega a\hat{d}'_{x,t} = \sum_{\text{all } x,t} 2\omega a^2 \hat{d}_{x,t}. \tag{16}$$

As a result,

$$updated(\hat{\theta}) = \hat{\theta} + \frac{\sum_{\text{all } x,t} 2\omega a(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega a^2 \hat{d}_{x,t}}. \tag{17}$$

The iterative procedure goes as follows:

1. set starting values for $\hat{\alpha}_x = \frac{1}{T}\sum_{\text{all t}} m_{x,t}$ (SVD estimate), $\hat{\beta}_x = 1$, $\hat{\kappa}_t = 0$, compute $\hat{d}_{x,t}$, $D(d_{x,t}, \hat{d}_{x,t})$;
2. update $\hat{\alpha}_x = \hat{\alpha}_x + \dfrac{\sum_{\text{all } x,t} 2\omega(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega\hat{d}_{x,t}}$, compute $\hat{d}_{x,t}$;
3. update $\hat{\kappa}_t = \hat{\kappa}_t + \dfrac{\sum_{\text{all } x,t} 2\omega\hat{\beta}_x(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega\hat{\beta}_x^2\hat{d}_{x,t}}$, adjust such that $\hat{\kappa}_t = \hat{\kappa}_t - \bar{\hat{\kappa}}_t$, compute $\hat{d}_{x,t}$;

4. update $\hat{\beta}_x = \hat{\beta}_x + \frac{\sum_{\text{all } x,t} 2\omega\hat{\kappa}_t(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega\hat{\kappa}_t^2\hat{d}_{x,t}}$, compute $\hat{d}_{x,t}$, $D_u(d_{x,t}, \hat{d}_{x,t})$;
5. check $\Delta D = D - D_u$.
If $\Delta D > 1 \times 10^{-6}$, replace D in step 1 by D_u and go to step 2;
Stop iterative process once $\Delta D \approx 0$ and take the fitted parameters as the ML estimates to the observed data;
Alternatively, stop if $\Delta D < 0$ for a consecutive 5 updating cycles and consider using other starting values or declare the iterations non-convergent.
6. once convergence is achieved, re-scale $\hat{\beta}$ and $\hat{\kappa}$ such that

$$\hat{\hat{\beta}} = \frac{\hat{\beta}}{\sum_{\text{all } x} \hat{\beta}_x}, \ \hat{\hat{\kappa}} = \hat{\kappa} \times (\sum_{\text{all } x} \hat{\beta}_x). \tag{18}$$

This generalized modelling methodology is implemented in the R package **ilc**. Figure 6 shows the plot of the parameter estimates against calendar year and Figure 7 displays the difference between the actual total deaths and the expected total deaths for each year. There are still curved patterns in the plot, however, the fits have been improved, as the majority of the residuals vary between -5000 and 5000, instead of between -10000 and 10000, and are more centered around the zero-line.

Figure 6: Plots of PB modelling parameter estimates against year.

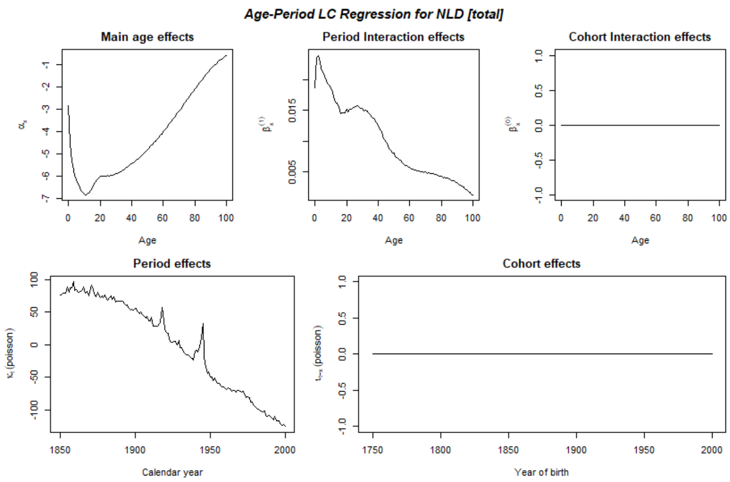
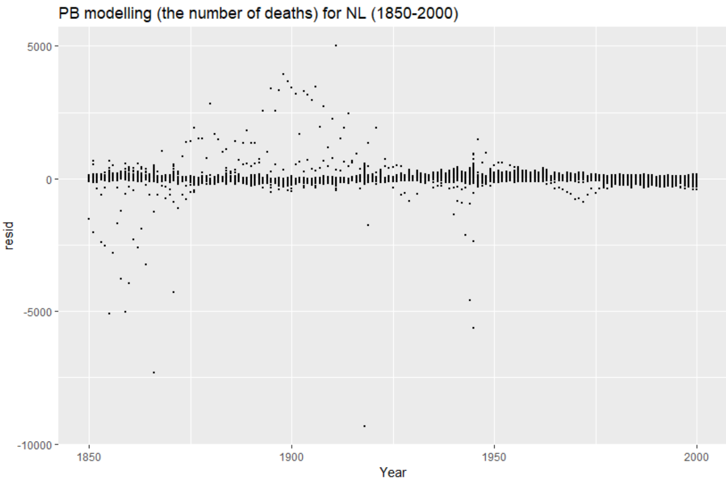


Figure 7: Residual plots against year.



3.2 The age-period-cohort Lee-Carter model

The model (1) can be further extended to include an additional bilinear term (as in Renshaw & Haberman (2006)). It is as follows:

$$ln(m_{x,t}) = \alpha_x + \beta_x^{(0)}\iota_{t-x} + \beta_x^{(1)}\kappa_t + \epsilon_{x,t},\ x \in [0, ..., X],\ t \in [1, .., T], \tag{19}$$

where ι_{t-x} represents the cohort effects and $\beta_x^{(0)}$ measures the interactions with age. To insure the identifiability of this model, (Renshaw & Haberman, 2008) impose the following constraints[1]

$$\sum_{x=0}^{X} \beta_x^{(0)} = 1, \sum_{x=0}^{X} \beta_x^{(1)} = 1, \text{ and } \iota_{1-X} = 0, \kappa_1 = 0. \tag{20}$$

The iterative procedure is slightly different from the Lee-Carter fitting procedure, as $\hat{\alpha}_x$ is estimated as $\frac{1}{T}\sum_{\text{all t}} m_{x,t}$ (SVD estimate) and not adjusted during the iterative process. Below shows the procedure:

1. Estimate $\hat{\alpha}_x = \frac{1}{T}\sum_{\text{all t}} m_{x,t}$ (SVD estimate);
2. Set starting values for $\hat{\beta}_x^{(0)} = \hat{\beta}_x^{(1)} = 1$, estimate the simplified period-cohort predictor: $ln(m_{x,t}) = ln(\epsilon_{x,t}) + \alpha_x + \iota_{t-x} + \kappa_t$, get the initial values for ι_{t-x} and κ_t, then compute $\hat{d}_{x,t}$, $D(d_{x,t}, \hat{d}_{x,t})$;
3. update $\hat{\iota}_{t-x} = \hat{\iota}_{t-x} + \dfrac{\sum_{\text{all } x,t} 2\omega\hat{\beta}_x^{(0)}(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega(\hat{\beta}_x^{(0)})^2\hat{d}_{x,t}}$, adjust such that $\hat{\iota}_{t-x} = \hat{\iota}_{t-x} - \hat{\iota}_{1-X}$, compute $\hat{d}_{x,t}$;
4. update $\hat{\beta}_x^{(0)} = \hat{\beta}_x^{(0)} + \dfrac{\sum_{\text{all } x,t} 2\omega\hat{\iota}_{t-x}(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega\hat{\iota}_{t-x}^2\hat{d}_{x,t}}$, compute $\hat{d}_{x,t}$;
5. update $\hat{\kappa}_t = \hat{\kappa}_t + \dfrac{\sum_{\text{all } x,t} 2\omega\hat{\beta}_x^{(1)}(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega(\hat{\beta}_x^{(1)})^2\hat{d}_{x,t}}$, adjust such that $\hat{\kappa}_t = \hat{\kappa}_t - \hat{\kappa}_1$, compute $\hat{d}_{x,t}$;
6. update $\hat{\beta}_x^{(1)} = \hat{\beta}_x^{(1)} + \dfrac{\sum_{\text{all } x,t} 2\omega\hat{\kappa}_t(d_{x,t} - \hat{d}_{x,t})}{\sum_{\text{all } x,t} 2\omega\hat{\kappa}_t^2\hat{d}_{x,t}}$, compute $\hat{d}_{x,t}$, $D_u(d_{x,t}, \hat{d}_{x,t})$;
7. check $\Delta D = D - D_u$.

[1] However, these constraints are not sufficient, which will be illustrated in later section.

If $\Delta D > 1 \times 10^{-6}$, replace D in step 2 by D_u and go to step 3;
Stop iterative process once $\Delta D \approx 0$ and take the fitted parameters as the ML estimates to the observed data;
Alternatively, stop if $\Delta D < 0$ for a consecutive 5 updating cycles and consider using other starting values or declare the iterations non-convergent.
8. once convergence is achieved, re-scale $\hat{\beta}^{(0)}$, $\hat{\beta}^{(1)}$, $\hat{\iota}$ and $\hat{\kappa}$ such that

$$\hat{\hat{\beta}}^{(0)} = \frac{\hat{\beta}^{(0)}}{\sum_{\text{all } x} \hat{\beta}_x^{(0)}}, \; \hat{\hat{\beta}}_x^{(1)} = \frac{\hat{\beta}^{(1)}}{\sum_{\text{all } x} \hat{\beta}^{(1)}}, \; \hat{\hat{\iota}} = \hat{\iota} \times (\sum_{\text{all } x} \hat{\beta}_x^{(0)}), \; \hat{\hat{\kappa}}_t = \hat{\kappa}_t \times (\sum_{\text{all } x} \hat{\beta}_x^{(1)}). \tag{21}$$

This generalized modelling methodology is implemented in the R package **ilc** as well. Figure 8 shows the plot of the parameter estimates against calendar year and Figure 9 the residual plots against calendar year. Clearly, the fits have been further improved, which is confirmed by the MSE calculated for the Lee-Carter model, PB modelling and age-period-cohort model, shown in Table 3.

Figure 8: Plots of age-period-cohort Lee-Carter parameter estimates against year.

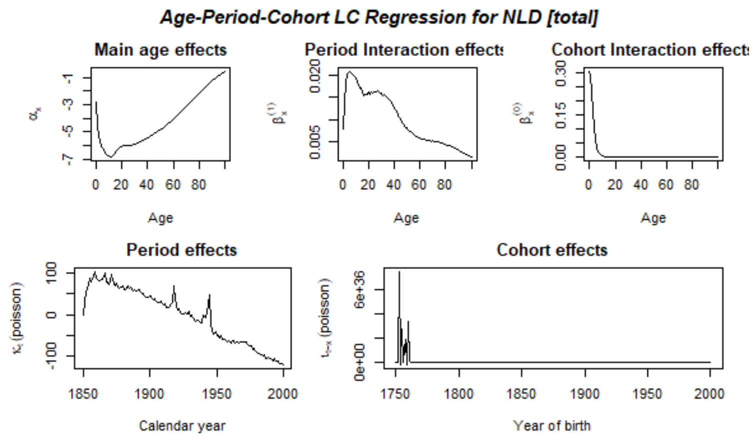


Figure 9: Residual plots against year.

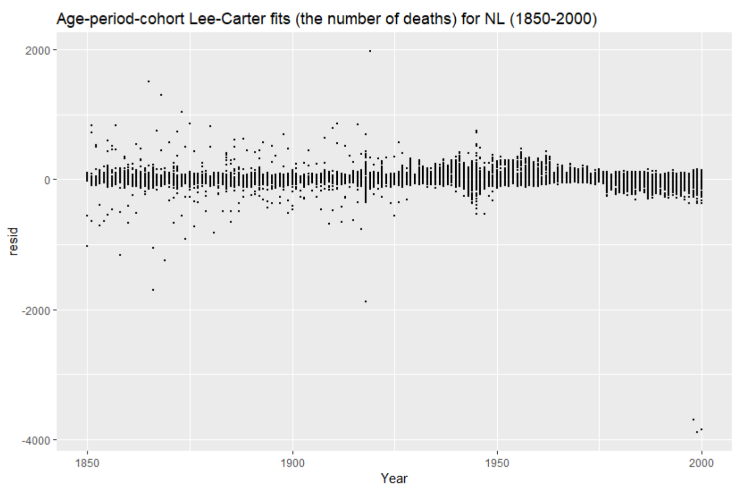
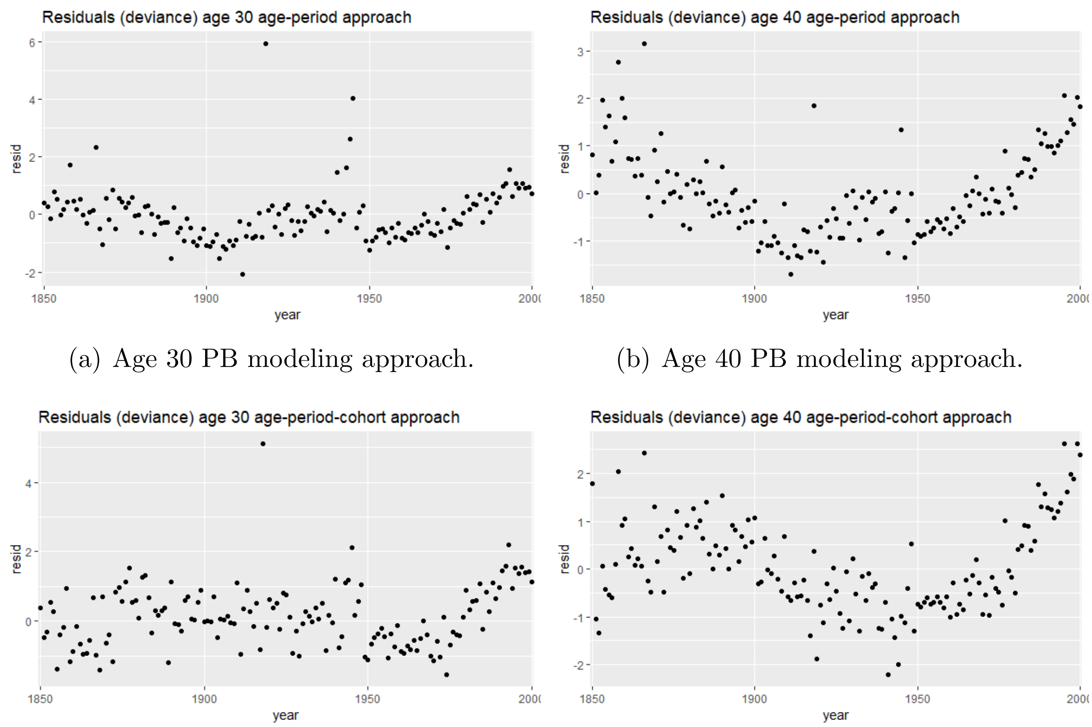


Table 3: MSE (w.r.t. deaths) of Lee-Carter, PB modelling and age-period-cohort models.

	Lee-Carter	PB modelling	age-period-cohort
MSE	206341	48266.3	11192.9

Moreover, the residual(deviance) plots against calender year for the age 30 and 40 are reproduced for PB modeling and age-period-cohort Lee-Carter models in Figure 10. To point out, deviance is defined as $sign(d_{x,t} - \hat{d}_{x,t})\sqrt{\frac{dev_{x,t}}{\hat{\psi}}}$ ($\hat{\psi} = \frac{D(d_{x,t}-\hat{d}_{x,t})}{v}$), where $\hat{\phi}$ is an empirical scaling factor and v represents the degrees of freedom, dependent on the particular model structure (Butt et al., 2015). The systematic patterns in Figure 10(a) and 10(b) imply that the model structures are likely not sufficiently flexible to represent adequately all age-specific differential trends in the data set. Figure 10(c) and 10(d) indicate a marked improvement in the goodness of fit by including the cohort effects.

Figure 10: Residual(deviance) plots.



(a) Age 30 PB modeling approach. (b) Age 40 PB modeling approach.

(c) Age 30 age-period-cohort approach. (d) Age 40 age-period-cohort approach.

3.2.1 Identifiability issue

Mentioned earlier that the imposed constraints (20) cannot insure the model identifiability. This can be illustrated with an example. Let $X > 2$, $T > 2$, then $1 - X < 0$. Take $(\alpha,\ \beta^{(1)},\ \kappa)$ such that the constraints on $\beta^{(1)}$ and κ are fulfilled. Furthermore, let $(\tilde{\alpha},\ \tilde{\beta}^{(1)},\ \tilde{\kappa}) = (\alpha,\ \beta^{(1)},\ \kappa)$. Define

$$\beta_x^{(0)} = \begin{cases} 0.75 & \text{if x=0,} \\ 0.25 & \text{if x=1,} \\ 0 & \text{otherwise;} \end{cases} \quad \iota_z = \begin{cases} 1 & \text{if z=0 or T,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\tilde{\beta}_x^{(0)} = \begin{cases} 0.5 & \text{if x=0,} \\ 0.5 & \text{if x=1,} \\ 0 & \text{otherwise;} \end{cases} \quad \tilde{\iota}_z = \begin{cases} 0.5 & \text{if z=0 ,} \\ 1.5 & \text{if z=T,} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\beta_x^{(0)}\iota_{t-x} = \tilde{\beta}_x^{(0)}\tilde{\iota}_{t-x} = \begin{cases} 0.25 & \text{if x=1, t=1,} \\ 0.75 & \text{if x=0, t=T,} \\ 0 & \text{otherwise.} \end{cases}$$

4 Inference pitfalls in Lee-Carter model for forecasting mortality

The paper by Leng & Peng (2016) states that mortality forecast and conclusions based on the two-step inference procedure for the Lee-Carter model could be inaccurate and questionable. To illustrate possible pitfalls behind the two-step estimation procedure, take one simpler Lee-Carter model considered by Leng & Peng (2016):

$$ln(m_{x,t}) = \alpha_x + \frac{1}{K}\kappa_t + \epsilon_{x,t},\ \kappa_t = \phi_0 + \phi_1\kappa_{t-1} + e_t \tag{22}$$

for x=1, ..., K and t=1, ..., T, where $\epsilon'_{x,t}$s and e'_ts are independent random errors with

$$E(\epsilon_{x,t}) = 0,\ E(\epsilon_{x,t}^2) = \sigma_x^2,\ E(e_t^2) = \sigma^2, \text{and constraint} \sum_t^T \kappa_t = 0. \tag{23}$$

Following the two-step estimation procedure in Lee & Carter (1992),

$$\hat{\alpha}_x = \frac{1}{T}\sum_{t=1}^{T} ln(m_{x,t}),\ \hat{\kappa}_t = \sum_{x=1}^{K}(ln(m_{x,t}) - \hat{\alpha}_x) \tag{24}$$

for x=1, ..., K and t=1, ..., T. ϕ_0, ϕ_1 are estimated by minimizing the following least squares

$$\sum_{t=1}^{T}(\hat{\kappa}_t - \phi_0 - \phi_1\hat{\kappa}_{t-1})^2. \tag{25}$$

then the least squares estimators for ϕ_1 based on (22) is:

$$\tilde{\phi}_1 = \frac{\sum_{t=1}^{T}\hat{\kappa}_t\sum_{s=1}^{T}\hat{\kappa}_{s-1} - T\sum_{t=1}^{t}\hat{\kappa}_t\hat{\kappa}_{t-1}}{(\sum_{t=1}^{T}\hat{\kappa}_{t-1})^2 - T\sum_{t=1}^{t}\hat{\kappa}_{t-1}^2} \tag{26}$$

Theorem Suppose model (22) holds with $\hat{\kappa}_0 = 0$, where $\{\epsilon_{x,t}\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d) random variables with zero mean and variance σ_x^2 for each x=1, ..., K, $\{e_t\}_{t=1}^T$ is sequence of i.i.d. random variables with zero mean and variance $\sigma^2 > 0$, $\{\epsilon_{x,t}\}_{t=1}^T$ for each x=1, ..., K and $\{e_t\}_{t=1}^T$ are independent of each other, and $\sum_{x=1}^T E|\epsilon_{x,1}|^\delta + E|e_1|^\delta < \infty$ for some $\delta > 2$.
(i) If $|\phi_1| < 1$ and ϕ_1 is independent of T, i.e., $\{\kappa_t\}$ is stationary, then $\tilde{\phi}_1 - \phi_1 \xrightarrow{p} \dfrac{-\phi_1\sigma_*^2}{\sigma_*^2 + \sigma^2/(1-\phi_1^2)}$ as T $\to \infty$, where $\sigma_*^2 = \sum_{x=1}^K \sigma_x^2$;
(ii) If $\phi_1 = 1 - \gamma/T$ for constant $\gamma \in R$, i.e. $\{\kappa_t\}$ is near unit root ($\gamma \neq 0$) and unit root ($\gamma = 0$), then $\tilde{\phi}_1 - \phi_1 \xrightarrow{p} 0$ as $T \to \infty$;

(iii) If $|\phi_1| > 1$ and ϕ_1 is independent of T, i.e., $\{\kappa\}$ is an explosive AR(1) process, then $\tilde{\phi}_1 - \phi_1 \xrightarrow{p} 0$ as $T \to \infty$.
The paper by Leng & Peng (2016) states that, it follows from this theorem that the two-step estimation in Lee & Carter (1992) does detect the true dynamics of κ_t's when $\{\kappa_t\}$ follows from an AR(1) process and the sequence is non-stationary. However, in case $\kappa_t = \phi_0 + \phi_1\kappa_{t-1} + \phi_2\kappa_{t-2} + e_t$, the estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ are inconsistent in general except for some special cases. Even when $\phi_1 \neq 0$ and $\phi_2 = 0$, they are inconsistent. Similar inconsistency is conjectured to hold for the general Lee-Carter model when $\{\kappa_t\}$ follows from an ARIMA(p,d,q) model with $p + d + q > 1$.

4.1 Simulation Study

This section will examine by simulation whether the two-step Lee-Carter estimation procedure can detect the true dynamics of κ_t's, given $\{\kappa_t\}$ follows from the ARIMA(0,1,0) model. Draw 10000 random samples from model (22) with K=24, T=100 or 1000 or 10000. Take $\phi_0 = -0.3$, $\phi_1 = 1$, $\alpha_x = 1$, $\sigma = 1.75$, and $\sigma_x = 0$ or 1 or 10 for x=1, ..., K. First follow Lee-Carter's two-step procedure to calculate $\hat{\kappa}_t$'s (without re-estimation). Then use equation (26) to obtain $\tilde{\phi}_1$. The results are displayed in Table 4. Additionally, use **auto.arima** to pick up the best ARIMA(p,d,0) by BIC with $p \leq 5$ and $d \leq 1$. The results are shown in Table 5.

Table 4: Mean and standard deviation for estimator $\tilde{\phi}_1$.

σ_x	T=100		T=1000		T=10000	
	mean	sd	mean	sd	mean	sd
0	0.9733	0.0328	0.9998	0.0007		
1	0.6980	0.2296	0.9998	0.0007		
10	-0.0051	0.1005	0.7449	0.0838	0.9968	0.0004

[1] $\phi_0 = -0.3, \phi_1 = 1, K = 24, \alpha_1 = ... = \alpha_K = 1, \sigma = 1.75$ in model (22), $N = 10000$.

Table 5: Best fit determined by auto.arima in the R package 'forecast'.

	$\sigma_x = 0$	$\sigma_x = 0.1$	$\sigma_x = 1$	$\sigma_x = 10$
ARIMA(1,0,0)	158	76	0	0
ARIMA(2,0,0)	3	3	0	0
ARIMA(3,0,0)	5	5	0	0
ARIMA(4,0,0)	12	4	0	0
ARIMA(5,0,0)	10	8	0	37
ARIMA(0,1,0)	4258	3379	0	0
ARIMA(1,1,0)	1472	2471	0	0
ARIMA(2,1,0)	1055	1040	0	0
ARIMA(3,1,0)	981	993	20	0
ARIMA(4,1,0)	912	856	2120	1
ARIMA(5,1,0)	1134	1165	7860	9962

[1] max.p=5, max.d=1, max.q=0 for model(22) with $\phi_0 = -0.3, \phi_1 = 1, K = 24, \alpha_1 = ... = \alpha_K = 1, \sigma = 1.75, T = 1000, N = 10000$.

Table 4 shows that, the two-step estimation in Lee & Carter (1992) can detect the true dynamics of κ_t's if T is large enough. Furthermore, Table 5 indicates that using $\hat{\kappa}_t$ and **auto.arima** fails to obtain the right model with this data. Even when $\sigma_x = 0$, the chance that ARIMA(0,1,0) is picked is low (42.58%).

5 The plug-in Lee-Carter models

Another issue with the Lee-Carter model is that the constraint $\sum_{t=1}^{T} \kappa_t = 0$ becomes a constraint on the possible realization of the stochastic process $\{\kappa_t\}$. Suppose $\sum_{t=1}^{T} \kappa_t = 0$, if increase T, κ_{T+1} or any κ_t with $t > T+1$ is expected to be 0, because the realization of $\kappa_1, ..., \kappa_T$ cannot be changed. This issue is pointed out by the paper Beutner, Reese & Urbain (2017), described as inconsistency arising from the dynamic view on the constraints. To avoid this implausible constraints issue, it suggests replacing κ_t by time series models from the beginning.

5.1 The plug-in age-period Lee-Carter model

The plug-in Lee-Carter model is as follows :

$$ln(m_{x,t}) = \alpha_x + \beta_x \mu t + \beta_x c + \beta_x \sum_{l=1}^{t} e_l + \epsilon_{x,t}, \quad x = 0, \cdots, X, \; t = 1, \cdots, T, \tag{27}$$

where κ_t is assumed to follow a random walk process such that $\kappa_t = \mu + \kappa_{t-1} + e_t = \mu t + c + \sum_{l=1}^{t} e_t, \;\; t \geq 1,$ with $\kappa_0 = c, \; c \in R$. However, as the mortality index is unobservable, whether the mortality index really follows from a random walk process cannot be tested. Furthermore, $\{e_t\}$ is a sequence of independent and identically distributed random variables with $E(e_t) = 0$ which is independent of $\epsilon_{x,t}$.
The parameter set is $\Theta := \big\{(\alpha, \beta, \mu, \sigma_\epsilon^2, \sigma_e^2) \in R^{X+1} \times R^{X+1} \times R \times R_+ \times R_+ | \sum_{x=0}^{X} \beta_x = 1\big\}$, where $R_+ = \{x \in R | x > 0\}$. The expected values and the covariance are given by:

$$E_\theta\big(ln(m_{x,t})\big) = \alpha_x + \beta_x \mu t + \beta_x c,$$
$$Cov_\theta\big(ln(m_{x,s}), ln(m_{y,t})\big) = \beta_x \beta_y \sigma_e^2 min\{s,t\} + 1_{x=y,s=t}(x,y,s,t)\sigma_\epsilon^2.$$

The expected values and the covariance structure are proved identifiable if $T \geq 2$, for more details refer to the paper by Beutner, Reese & Urbain (2017).

5.1.1 LS estimation

To estimate the parameters $(\alpha, \beta, \mu, \sigma_\epsilon^2, \sigma_e^2)$, the paper by Beutner, Reese & Urbain (2017) suggests minimizing the following objective function:

$$\sum_{x=0}^{X} \sum_{t=1}^{T} \frac{1}{t} \big(ln(m_{x,t}) - (\alpha_x + \beta_x \mu t)\big)^2$$
$$+ \sum_{x=0}^{X} \sum_{s=1}^{T} \sum_{y=0}^{X} \sum_{t=1}^{T} \big(min\{s,t\}\big)^{-1} \Bigg(ln(m_{x,s}) ln(m_{y,t})$$
$$- \left[\beta_x \beta_y \sigma_e^2 min\{s,t\} + 1_{\{x=y,s=t\}} \sigma_\epsilon^2 + (\alpha_x + \beta_x \mu s)(\alpha_y + \beta_y \mu t)\right] \Bigg)^2.$$

Note that c is set equal to 0. Moreover, the function **solnp** of the R package **Rsolnp** can be used to estimate the parameters. For illustrative purposes, apply this approach to the Dutch life table 5x1, 1850-2016 inclusive, with age grouped and classified as {<1, 1-4, 5-9, 10-14, ... , 105-109,110+}. The data are retrieved from the website https://www.mortality.org. The first five observations are shown in Table 6.

Table 6: First five observations of the Dutch life table 5x1.

Year	age	mx	expo	deaths
1850	0	0.22385	87284.26	19539
1850	1-4	0.03393	264251.2	9017
1850	5-9	0.00854	345254.8	2950
1850	10-14	0.00477	330880.6	1581
1850	15-19	0.00582	285955	1653

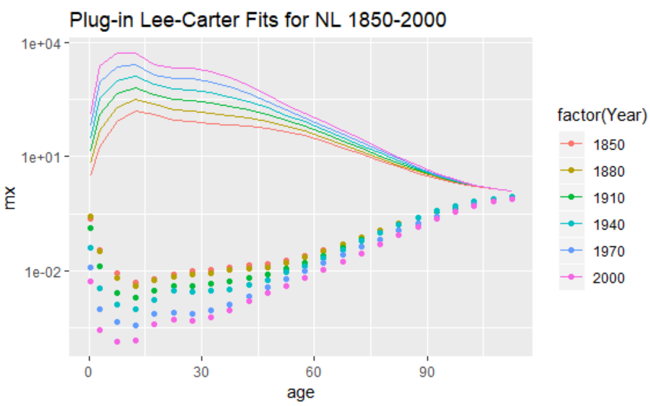
[1] mx: mortality rates, expo: exposure to risk.

Divide it into two subsets such that the training set contains observations before the year 2001 and the test set contains the rest observations. The estimates obtained are displayed in Table 7 and the fitted mortality for several years between 1850 and 2000 in Figure 11. Clearly, the plug-in age-period Lee-Carter model with estimates obtained using this objective function doesnot fit the data.

Table 7: Estimates plug-in age-period Lee-Carter model.

$\hat{\alpha}_0$	1.105531	$\hat{\alpha}_{12}$	3.585316	$\hat{\beta}_0$	0.083616	$\hat{\beta}_{12}$	0.030326	$\hat{\mu}$	0.298604075
$\hat{\alpha}_1$	2.891789	$\hat{\alpha}_{13}$	3.241471	$\hat{\beta}_1$	0.108418	$\hat{\beta}_{13}$	0.026224	$\hat{\sigma}^2_\epsilon$	0.041667
$\hat{\alpha}_2$	4.397282	$\hat{\alpha}_{14}$	2.819621	$\hat{\beta}_2$	0.092939	$\hat{\beta}_{14}$	0.023995	$\hat{\sigma}^2_e$	0.041667
$\hat{\alpha}_3$	5.019605	$\hat{\alpha}_{15}$	2.436379	$\hat{\beta}_3$	0.079678	$\hat{\beta}_{15}$	0.020545		
$\hat{\alpha}_4$	4.840658	$\hat{\alpha}_{16}$	2.035744	$\hat{\beta}_4$	0.066946	$\hat{\beta}_{16}$	0.017439		
$\hat{\alpha}_5$	4.492329	$\hat{\alpha}_{17}$	1.664954	$\hat{\beta}_5$	0.069936	$\hat{\beta}_{17}$	0.013732		
$\hat{\alpha}_6$	4.395389	$\hat{\alpha}_{18}$	1.299381	$\hat{\beta}_6$	0.072263	$\hat{\beta}_{18}$	0.009713		
$\hat{\alpha}_7$	4.28776	$\hat{\alpha}_{19}$	0.980001	$\hat{\beta}_7$	0.070297	$\hat{\beta}_{19}$	0.006967		
$\hat{\alpha}_8$	4.173877	$\hat{\alpha}_{20}$	0.729221	$\hat{\beta}_8$	0.065065	$\hat{\beta}_{20}$	0.004152		
$\hat{\alpha}_9$	4.109302	$\hat{\alpha}_{21}$	0.528192	$\hat{\beta}_9$	0.055543	$\hat{\beta}_{21}$	0.001191		
$\hat{\alpha}_{10}$	4.027725	$\hat{\alpha}_{22}$	0.357691	$\hat{\beta}_{10}$	0.04538	$\hat{\beta}_{22}$	0.000111		
$\hat{\alpha}_{11}$	3.814963	$\hat{\alpha}_{23}$	0.250923	$\hat{\beta}_{11}$	0.036344	$\hat{\beta}_{23}$	-0.00082		

Figure 11: Plot of the fitted value of $m_{x,t} : e^{\hat{\alpha}_x+\hat{\beta}_x\hat{\mu}t}$ for NL 5x1 1850-2000.

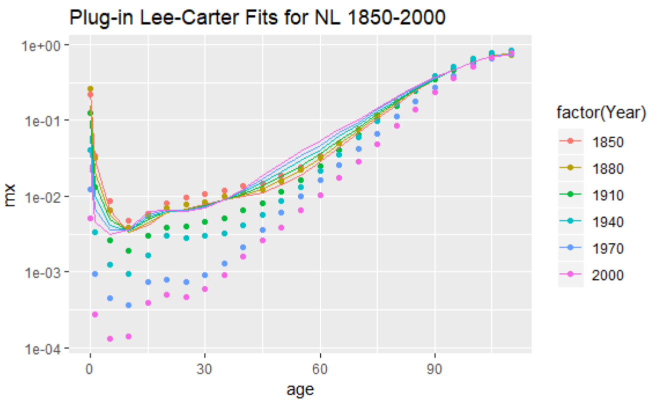


Adjust the objective function to the objective function shown below:

$$\sum_{x=0}^{X}\sum_{t=1}^{T}\frac{1}{t}\big(ln(m_{x,t})-(\alpha_x+\beta_x\mu t)\big)^2.$$

The estimated α, β, μ are displayed in Table 11 in the Appendix and the fitted mortality for these years between 1850 and 2000 in Figure 12.

Figure 12: Plot of the fitted value of $m_{x,t}$: $e^{\hat{\alpha}_x+\hat{\beta}_x\hat{\mu}t}$ for NL 5x1 1850-2000.

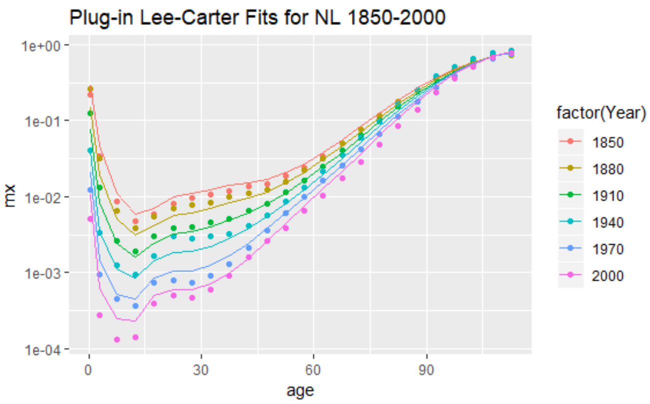


The results are further improved if include c in the objective function:

$$\sum_{x=0}^{X}\sum_{t=1}^{T}\frac{1}{t}\big(ln(m_{x,t})-(\alpha_x+\beta_x\mu t+\beta_x c)\big)^2.$$

The estimates are shown in Table 8 and the plot of the fitted mortality for the same years are displayed in Figure 13.

Figure 13: Plot of the fitted value of $m_{x,t}$: $e^{\hat{\alpha}_x+\hat{\beta}_x\hat{\mu}t+\hat{\beta}_x\hat{c}}$ for NL 5x1 1850-2000.



Then based on these estimates, estimate σ_e^2 and σ_ϵ^2 by using the objective function:

$$\sum_{x=0}^{X}\sum_{s=1}^{T}\sum_{y=0}^{X}\sum_{t=1}^{T}min\{s,t\}^{-1}\Big(ln(m_{x,s})ln(m_{y,t})$$
$$-[\hat{\beta}_x\hat{\beta}_y\sigma_e^2 min\{s,t\}+1_{\{x=y,s=t\}}\sigma_\epsilon^2+(\hat{\alpha}_x+\hat{\beta}_x\hat{\mu}s+\hat{\beta}_x\hat{c})(\hat{\alpha}_y+\hat{\beta}_y\hat{\mu}t+\hat{\beta}_y\hat{c})]\Big)^2$$

Table 8: Estimates plug-in approach with objective function adjusted.

$\hat{\alpha}_0$	-0.42182	$\hat{\alpha}_{12}$	-3.31092	$\hat{\beta}_0$	0.081241	$\hat{\beta}_{12}$	0.031215	$\hat{\mu}$	-0.26998
$\hat{\alpha}_1$	-1.98466	$\hat{\alpha}_{13}$	-3.00064	$\hat{\beta}_1$	0.106913	$\hat{\beta}_{13}$	0.026906	$\hat{c}$	-9.76762
$\hat{\alpha}_2$	-3.56261	$\hat{\alpha}_{14}$	-2.64215	$\hat{\beta}_2$	0.094055	$\hat{\beta}_{14}$	0.02358	$\hat{\sigma}^2_\epsilon$	0.034937
$\hat{\alpha}_3$	-4.33257	$\hat{\alpha}_{15}$	-2.25626	$\hat{\beta}_3$	0.079791	$\hat{\beta}_{15}$	0.020669	$\hat{\sigma}^2_e$	0.000754
$\hat{\alpha}_4$	-4.29741	$\hat{\alpha}_{16}$	-1.8908	$\hat{\beta}_4$	0.06537	$\hat{\beta}_{16}$	0.017084		
$\hat{\alpha}_5$	-3.91129	$\hat{\alpha}_{17}$	-1.55253	$\hat{\beta}_5$	0.069241	$\hat{\beta}_{17}$	0.013257		
$\hat{\alpha}_6$	-3.76842	$\hat{\alpha}_{18}$	-1.21216	$\hat{\beta}_6$	0.07249	$\hat{\beta}_{18}$	0.010255		
$\hat{\alpha}_7$	-3.67256	$\hat{\alpha}_{19}$	-0.94708	$\hat{\beta}_7$	0.070922	$\hat{\beta}_{19}$	0.006316		
$\hat{\alpha}_8$	-3.59611	$\hat{\alpha}_{20}$	-0.70991	$\hat{\beta}_8$	0.065875	$\hat{\beta}_{20}$	0.00346		
$\hat{\alpha}_9$	-3.60859	$\hat{\alpha}_{21}$	-0.51676	$\hat{\beta}_9$	0.05669	$\hat{\beta}_{21}$	0.001322		
$\hat{\alpha}_{10}$	-3.63412	$\hat{\alpha}_{22}$	-0.36553	$\hat{\beta}_{10}$	0.045684	$\hat{\beta}_{22}$	-5.06E-05		
$\hat{\alpha}_{11}$	-3.48616	$\hat{\alpha}_{23}$	-0.26518	$\hat{\beta}_{11}$	0.038396	$\hat{\beta}_{23}$	-0.00068		

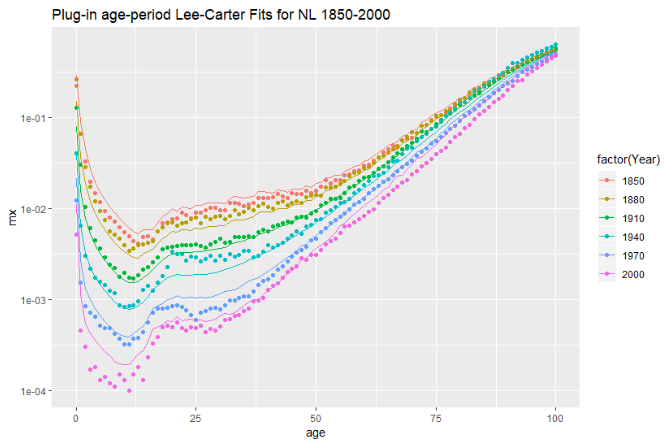
It is worth mentioning that the function **solnp** requires starting values, and different starting values result in different estimates. The earlier estimates are generated with starting value $\frac{1}{X+1}$ for all estimators. Table 9 shows the estimates generated with starting value 0.5 for all estimators. Very likely, starting values have impact on the forecasts.

Table 9: Estimates plug-in approach with objective function adjusted.

$\hat{\alpha}_0$	-1.60362	$\hat{\alpha}_{12}$	-3.76531	$\hat{\beta}_0$	0.081263	$\hat{\beta}_{12}$	0.031203	$\hat{\mu}$	-0.27005
$\hat{\alpha}_1$	-3.54099	$\hat{\alpha}_{13}$	-3.39236	$\hat{\beta}_1$	0.106863	$\hat{\beta}_{13}$	0.026908	$\hat{c}$	4.788814
$\hat{\alpha}_2$	-4.93169	$\hat{\alpha}_{14}$	-2.98559	$\hat{\beta}_2$	0.094013	$\hat{\beta}_{14}$	0.023563	σ^2_ϵ	0.03261
$\hat{\alpha}_3$	-5.49355	$\hat{\alpha}_{15}$	-2.5572	$\hat{\beta}_3$	0.079804	$\hat{\beta}_{15}$	0.020666	σ^2_e	0.0005
$\hat{\alpha}_4$	-5.24849	$\hat{\alpha}_{16}$	-2.13952	$\hat{\beta}_4$	0.065394	$\hat{\beta}_{16}$	0.017106		
$\hat{\alpha}_5$	-4.919	$\hat{\alpha}_{17}$	-1.74563	$\hat{\beta}_5$	0.069219	$\hat{\beta}_{17}$	0.013264		
$\hat{\alpha}_6$	-4.82346	$\hat{\alpha}_{18}$	-1.36152	$\hat{\beta}_6$	0.072476	$\hat{\beta}_{18}$	0.010285		
$\hat{\alpha}_7$	-4.70489	$\hat{\alpha}_{19}$	-1.03898	$\hat{\beta}_7$	0.07089	$\hat{\beta}_{19}$	0.006347		
$\hat{\alpha}_8$	-4.55491	$\hat{\alpha}_{20}$	-0.76029	$\hat{\beta}_8$	0.065843	$\hat{\beta}_{20}$	0.003488		
$\hat{\alpha}_9$	-4.43377	$\hat{\alpha}_{21}$	-0.53605	$\hat{\beta}_9$	0.056678	$\hat{\beta}_{21}$	0.001343		
$\hat{\alpha}_{10}$	-4.29913	$\hat{\alpha}_{22}$	-0.36491	$\hat{\beta}_{10}$	0.045675	$\hat{\beta}_{22}$	-2.18E-05		
$\hat{\alpha}_{11}$	-4.04508	$\hat{\alpha}_{23}$	-0.25537	$\hat{\beta}_{11}$	0.03838	$\hat{\beta}_{23}$	-0.00065		

Apply the plug-in age-period Lee-Carter model with adjusted objective function to the Dutch life table 1x1. The estimates obtained are displayed in Table 12 in the Appendix. The MSE with respect to mortality rates calculated for the Lee-Carter and the plug-in Lee-Carter models are shown in table 10. It shows that Lee-Carter model fit the data a little better. The fitted mortality for several years between 1850 and 2000 are presented in Figure 14.

Figure 14: Plot of the fitted value of $m_{x,t} : e^{\hat{\alpha}_x + \hat{\beta}_x \hat{\mu} t + \hat{\beta}_x \hat{c}}$ for NL 1x1 1850-2000.



5.1.2 Forecasting with the plug-in age-period Lee-Carter Model

Definition of the best one-step predictor The best one-step predictor $\hat{Z}_{t+1}$ is the random vector whose i^{th} component is the best linear mean square predictor of Z_t in terms of all the components of $ln(m_0) := 1, ln(m_1), \ldots, ln(m_{t-1})$.
To do the forecasting, the paper by Beutner, Reese & Urbain (2017) suggests rewriting the plug-in Lee-Carter model as an state space model. It is as follows. Let $t \geq 1$,

$$ln(m_t) = \begin{pmatrix} ln(m_{x,t}) \\ \vdots \\ ln(m_{X,t}) \end{pmatrix}, \; Z_t = \begin{pmatrix} \kappa_t \\ 1 \end{pmatrix}, \; F^2 = \begin{pmatrix} 1 & \mu + c \\ 0 & 1 \end{pmatrix}, \; G = \begin{pmatrix} \beta_0 & \alpha_0 \\ \vdots & \vdots \\ \beta_X & \alpha_X \end{pmatrix} \tag{28}$$

and

$$\mathrm{e}_t = \begin{pmatrix} e_t \\ 0 \end{pmatrix}, \; \epsilon_t = \begin{pmatrix} \epsilon_{0,t} \\ \vdots \\ \epsilon_{X,t} \end{pmatrix}. \tag{29}$$

Then,

$$ln(m_t) = GZ_t + \epsilon_t, \; Z_{t+1} = FZ_t + \mathrm{e}_{t+1}, \tag{30}$$

where $\{\mathrm{e}_t\}$ is a sequence of random n-vectors uncorrelated with Z_t such that

$$E\epsilon_j\epsilon_k' = \begin{cases} R_k & \text{if j=k,} \\ 0 & \text{otherwise;} \end{cases} \quad E\mathrm{e}_j\mathrm{e}_k' = \begin{cases} Q_k & \text{if j=k,} \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

$\{R_k\}$ and $\{Q_k\}$ are a sequence of known covariance matrices. However R_k and Q_k are constant in this context,

$$\underset{(X+1)\times(X+1)}{R_k} = R = \begin{pmatrix} \sigma_\epsilon^2 & 0 & \ldots & 0 \\ 0 & \sigma_\epsilon^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_\epsilon^2 \end{pmatrix}, \; \underset{2\times2}{Q_k} = Q = \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{pmatrix}. \tag{32}$$

Additionally, it is assumed that

$$\epsilon_j \perp \mathrm{e}_k \text{ for all j,k; } E\epsilon_j Z_1' = E\mathrm{e}_j Z_1' = 0 \text{ for all j.} \tag{33}$$

Take the unconditional expectation of Z_1: $\hat{Z}_1 = \begin{pmatrix} \hat{\mu}+\hat{c} \\ 1 \end{pmatrix}$, as a starting value for the latent Z_1 and replace the unknown parameters by their estimates. Furthermore, the following quantities are needed.

$$\Psi_1 = E[\hat{Z}_1\hat{Z}_1'] = \begin{pmatrix} (\hat{\mu}+\hat{c})^2 & \hat{\mu}+\hat{c} \\ \hat{\mu}+\hat{c} & 1 \end{pmatrix}, \ \Pi_1 = E[Z_1Z_1'] = \begin{pmatrix} (\hat{\mu}+\hat{c})^2+\hat{\sigma_e}^2 & \hat{\mu}+\hat{c} \\ \hat{\mu}+\hat{c} & 1 \end{pmatrix}, \tag{34}$$

and

$$\Phi_1 = \Pi_1 - \Psi_1 = \begin{pmatrix} \hat{\sigma_e}^2 & \hat{\mu}+\hat{c} \\ \hat{\mu}+\hat{c} & 1 \end{pmatrix}. \tag{35}$$

Based on the observed values $ln(m_1),\ \ldots,\ ln(m_T)$, Z_{T+1} can be predicted using the initial values and the following recursion for $t = 1,\ldots,T$,

$$\Delta_t = \hat{G}\phi_t\hat{G}' + \hat{R},\ \Xi_t = \hat{F}\Phi_t\hat{G}',\ \hat{Z}_{t+1} = \hat{F}\hat{Z}_t + \Xi_t\Delta_t^{-1}(ln(m_t) - \hat{G}\hat{Z}_t), \tag{36}$$

and

$$\Pi_{t+1} = \hat{F}\Pi_t\hat{F}' + \hat{Q}_{t+1}, \Psi_{t+1} = \hat{F}\Psi_t\hat{F}' + \Xi_t\Delta_t'\Xi',\ \Phi_{t+1} = \Pi_{t+1} - \Psi_{t+1}, \tag{37}$$

where Δ_t^{-1} denotes a (generalized) inverse, $\hat{F}$ and $\hat{G}$ equal F and G with μ replaced by $\hat{\mu}$, and β_x, α_x, $x = 0,\ldots,X$, replaced by $\hat{\beta}_x$ and $\hat{\alpha}_x$, respectively[3]. It is worth mentioning that the second row of $\hat{Z}_{t+1}$ remains equal to 1.

Proof Let

$$v_t = ln(m_t) - G\hat{Z}_t. \tag{38}$$

Then

$$v_t \overset{(30)}{=} GZ_t + \epsilon_t - G\hat{Z}_t = G(Z_t - \hat{Z}_t) + \epsilon_t, \tag{39}$$

as $v_t \perp \hat{Z}_t$, $\epsilon_t \perp (Z_t - \hat{Z}_t)$,

$$E(v_tv_t') = G\Phi_tG' + R_t. \tag{40}$$

Furthermore,

$$E(Z_tv_t') = E[(Z_t - \hat{Z}_t)v_t'] = E[(Z_t - \hat{Z}_t)v_t'] \overset{(37)}{=} E[(Z_t - \hat{Z}_t)(G(Z_t - \hat{Z}_t) + \epsilon_t)'] = \Phi_tG'. \tag{41}$$

Let P denote 'projection', then the following equation holds for Z_t and $ln(m_t)$ defined as in equations (28) and (30).

$$\begin{aligned} \hat{Z}_{t+1} &= P\big(FZ_t + e_{t+1}|ln(m_0),\ldots,ln(m_t)\big) \\ &= P\big(FZ_t + e_{t+1}|ln(m_0),\ldots,ln(m_{t-1})\big) + P(FZ_t + e_{t+1}|v_t) \\ &= F\hat{Z}_t + K_tv_t, \end{aligned} \tag{42}$$

where $K_t = E(FZ_t + e_{t+1})v_t')[E(v_tv_t')]^{-1}$, given $E(v_tv_t')$ is non-singular. Since $v_t \perp \hat{Z}_t$,

$$\begin{aligned} K_t &\overset{(39)}{=} E(FZ_t + e_{t+1})(G(Z_t - \hat{Z}_t) + \epsilon_t)')[E(v_tv_t')]^{-1} \\ &= E(F(Z_t - \hat{Z}_t) + e_{t+1})(G(Z_t - \hat{Z}_t) + \epsilon_t)')[E(v_tv_t')]^{-1} \\ &= FE[(Z_t - \hat{Z}_t)(Z_t - \hat{Z}_t)']G'[E(v_tv_t')]^{-1} \\ &\overset{(40)}{=} F\Phi_tG'(G\Phi_tG' + R_t)^{-1}. \end{aligned} \tag{43}$$

[3] $\Pi_{t+1} = \hat{F}\Pi_t\hat{F}'$ in the paper by Beutner, Reese & Urbain (2017).

Moreover,

$$
\begin{aligned}
\Pi_{t+1} &= E[Z_{t+1}Z'_{t+1}] \\
&\overset{(30)}{=} E[(FZ_t + \mathrm{e}_{t+1})(FZ_t + \mathrm{e}_{t+1})'] \\
&= FE[Z_tZ'_t]F' + E[e_{t+1}e'_{t+1}] \\
&= F\Pi_tF' + Q, \\
\Psi_{t+1} &= E[\hat{Z}_{t+1}\hat{Z}'_{t+1}] \\
&= E[(F\hat{Z}_t + K_tv_t)(F\hat{Z}_t + K_tv_t)'] \\
&= F\Psi_tF' + K_tE[v_tv'_t]K'_t, \\
&= F\Psi_tF' + F\Phi_tG'(G\Phi_tG' + R_t)^{-1}(F\Phi_tG')', \\
\Phi_{t+1} &= \Pi_{t+1} - \Psi_{t+1} \\
&= F\Phi_tF' + Q - F\Phi_tG'(G\Phi_tG' + R_t)^{-1}(F\Phi_tG')'.
\end{aligned}
\tag{44}
$$

Shown by Brockwell & Davis (1987) section 12.1, the following equation holds:

$$
\Phi_{t+1} = Q + (F - K_tG)\Phi_t(F - K_tG)' + K_tR_tK'_t.
$$

However, these two Φ_{t+1}'s are equivalent, which can be shown as follows.

$$
\begin{aligned}
\Phi_{t+1} &= Q + (F - K_tG)\Phi_t(F - K_tG)' + K_tR_tK'_t \\
&= Q + F\Phi_tF' - F\Phi_tG'_tK'_t - K_tG\Phi_tF' + K_tG\Phi_t(K_tG)' + K_tR_tK'_t \\
&= Q + F\Phi_tF' - F\Phi_tG'_tK'_t - K_tG\Phi_tF' + K_t(G\Phi_tG' + R_t)K'_t \\
&\overset{(43)}{=} Q + F\Phi_tF' - F\Phi_tG'_tK'_t - K_tG\Phi_tF' \\
&\quad + F\Phi_tG'(G\Phi_tG' + R_t)^{-1}(G\Phi_tG' + R_t)(G\Phi_tG' + R_t)^{-1}(F\Phi_tG')' \\
&= Q + F\Phi_tF' - F\Phi_tG'_tK'_t - K_tG\Phi_tF' + F\Phi_tG'_tK'_t \\
&= Q + F\Phi_tF' - K_tG\Phi_tF' \\
&= Q + F\Phi_tF' - F\Phi_tG'(G\Phi_tG' + R_t)^{-1}(F\Phi_tG')'.
\end{aligned}
\tag{45}
$$

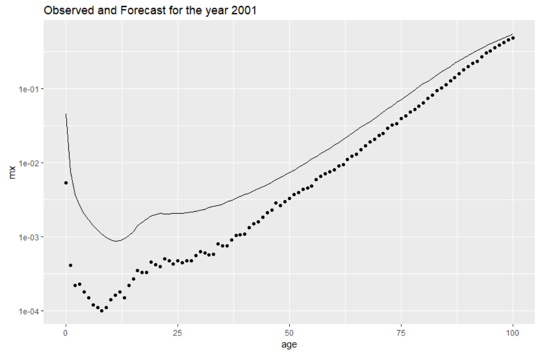
If replace F, G, Q and Z_t by $\hat{F}$, $\hat{G}$, $\hat{Q}$ and $\hat{Z}_t$, equations (36), (37) are proved.

As a result, $\widehat{ln(m_{t+1})} = \hat{G}\hat{Z}_{t+1}$, and $\widehat{ln(m_{t+h})} = \hat{G}\hat{F}^{h-1}\hat{Z}_{t+1}$, $h = 2, 3, \ldots$, with

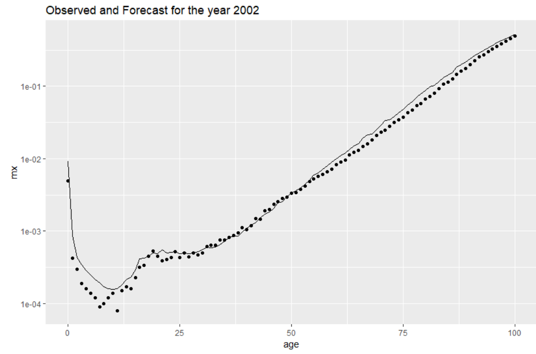
$$
\hat{F}^{h-1} = \begin{pmatrix} 1 & h\hat{\mu} + \hat{c} \\ 0 & 1 \end{pmatrix}.
$$

The forecasts for the year 2001, 2002, 2003 and 2016 are displayed in Figure 15.

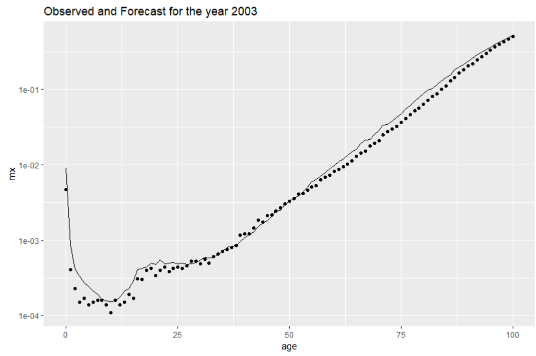
Figure 15: Plot of forecasting results.



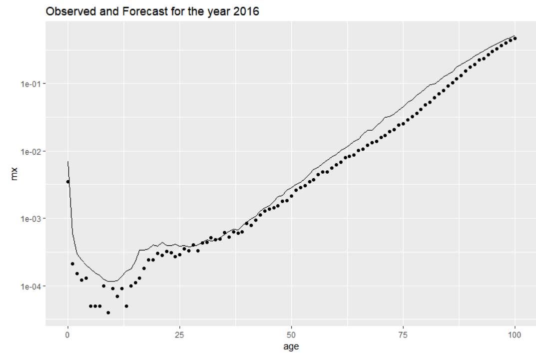
(a) Year 2001 NL.



(b) Year 2002 NL.



(c) Year 2003 NL.



(d) Year 2016 NL.

Seen from Figure 15, the plug-in Lee-Carter model generates proper forecasts for several years, such as the year 2002 and 2003, but extremely bad forecasts for the year 2001, which is one-year ahead forecasting. However, forecasting performance will be improved if critical intervals are included. Compared to Figure 4, the point forecasting performance of the plug-in age-period Lee-Carter model doesnot outperform the standard Lee-Carter model in case of this data.

5.2 The plug-in age-period-cohort Lee-Carter model

Recall that the age-period-cohort Lee-Carter model with proposed constraints (20) is not identifiable, however, Beutner, Reese & Urbain (2017) shows that the plug-in age-period-cohort model is identifiable. Let κ_t and ι_{t-x} be given by:

$$\begin{aligned} \kappa_t &= \kappa_0 + t\mu_1 + \sum_{s=1}^{t} e_s^{(1)}, \\ \iota_{t-x} &= \iota_{-X} + (t-x+X)\mu_0 + \sum_{t=1}^{t-x+X} e_{t-X}^{(0)}, \end{aligned} \tag{46}$$

where $\{e_t^{(0)}\}$ and $\{e_t^{(1)}\}$ are two independent sequences of independent and identically distributed random variables (i.i.d.) with expected values equal to zero and finite second moments denoted by

σ^2_{e0} and σ^2_{e1}, respectively. The plug-in age-period-cohort Lee-Carter model is given by:

$$ln(m_{x,t}) = \alpha_x+\beta_x^{(0)}c_0+\beta_x^{(0)}\mu_0+\beta_x^{(0)}\mu_0(t-x+X)+\beta_x^{(1)}c_1+\beta_x^{(1)}\mu_1 t+\beta_x^{(0)}\sum_{t=1}^{t-x+X} e^{(0)}_{t-X}+\beta_x^{(1)}\sum_{s=1}^{t} e^{(1)}_s+\epsilon_{x,t}. \tag{47}$$

Furthermore, $\{e_t^{(0)}\}$ and $\{e_t^{(1)}\}$ are independent of $\epsilon_{x,t}, x = 0, \dots, X, t = 1, \dots, T$.
The parameter set is $\Theta := \{(\alpha, \beta^{(0)}, \beta^{(1)}, \mu_0, \mu_1, \sigma^2_\epsilon, \sigma^2_{e0}, \sigma^2_{e1}) \in R^{X+1} \times R^{X+1} \times R^{X+1} \times R \times R \times R_+ \times R_+ \times R_+ | \sum_{x=0}^X \beta_x^{(i)} = 1, i = 0, 1, \beta^{(0)} \neq \beta^{(1)}\}$, where $R_+ = \{x \in R | x > 0\}$. The expected values and the covariance are given by:

$$\begin{aligned} E_\theta\big(ln(m_{x,t})\big) =& \alpha_x + \beta_x^{(0)}c_0 + \beta_x^{(0)}\mu_0 + \beta_x^{(0)}\mu_0(t-x+X) + \beta_x^{(1)}c_1 + \beta_x^{(1)}\mu_1 t, \\ Cov_\theta\big(ln(m_{x,s}), ln(m_{y,t})\big) =& \beta_x^{(0)}\beta_y^{(0)}\sigma^2_{e0} min\{s-x+X, t-y+X\} + \beta_x^{(1)}\beta_y^{(1)}\sigma^2_{e1} min\{s,t\} \\ &+ 1_{x=y,s=t}(x,y,s,t)\sigma^2_\epsilon. \end{aligned}$$

The expected values and the covariance structure are proved identifiable if $T > X + 2$ $(X > 0)$, for more details refer to the paper by Beutner, Reese & Urbain (2017).
Similarly, estimate parameters $\alpha, \beta^{(0)}, \beta^{(1)}, \mu_0, \mu_1, c_0, c_1$ by using the following objective function:

$$\sum_{x=0}^X \sum_{t=1}^T \frac{1}{t}\Big(ln(m_{x,t}) - (\alpha_x + \beta_x^{(0)}c_0 + \beta_x^{(0)}\mu_0 + \beta_x^{(0)}\mu_0(t-x+X) + \beta_x^{(1)}c_1 + \beta_x^{(1)}\mu_1 t)\Big)^2.$$

Then based on these estimates, estimate σ^2_{e0}, σ^2_{e1} and σ^2_ϵ by using the objective function:

$$\begin{aligned} &\sum_{x=0}^X \sum_{s=1}^T \sum_{y=0}^X \sum_{t=1}^T min\{s,t\}^{-1}\Big(ln(m_{x,s})ln(m_{y,t}) \\ &-\big[\hat\beta_x^{(0)}\hat\beta_y^{(0)}\sigma^2_{e0} min\{s-x+X, t-y+X\} + \hat\beta_x^{(1)}\hat\beta_y^{(1)}\sigma^2_{e1} min\{s,t\} + 1_{x=y,s=t}(x,y,s,t)\sigma^2_\epsilon \\ &\quad + \big(\hat\alpha_x + \hat\beta_x^{(0)}c_0 + \hat\beta_x^{(0)}\mu_0 + \hat\beta_x^{(0)}\hat\mu_0(s-x+X) + \hat\beta_x^{(1)}\hat c_1 + \hat\beta_x^{(1)}\hat\mu_1 s\big) \\ &\quad \times \big(\hat\alpha_y + \hat\beta_y^{(0)}c_0 + \hat\beta_y^{(0)}\hat\mu_0 + \hat\beta_y^{(0)}\hat\mu_0(t-y+X) + \hat\beta_y^{(1)}c_1 + \hat\beta_y^{(1)}\hat\mu_1 t\big)\big]\Big)^2. \end{aligned}$$

The estimates obtained are displayed in Table 13 in the Appendix and the fitted mortality for several years between 1850 and 2000 in Figure 16. There is no significant improvement compared with Figure 14, which is confirmed by the MSE calculated for both models in Table 10. Furthermore, forecasting with plug-in age-period-cohort Lee-Carter model is more complicated than with plug-in age-period Lee-Carter model and it is not considered here.

Figure 16: Plot of the fitted value of $m_{x,t}$: $e^{\hat{\alpha}_x+\hat{\beta}_x^{(0)}\hat{c}_0+\hat{\beta}_x^{(0)}\hat{\mu}_0(t-x-X)+\hat{\beta}_x^{(1)}\hat{c}_1+\hat{\beta}_x^{(1)}\hat{\mu}_1 t}$ for NL 1x1 1850-2000.

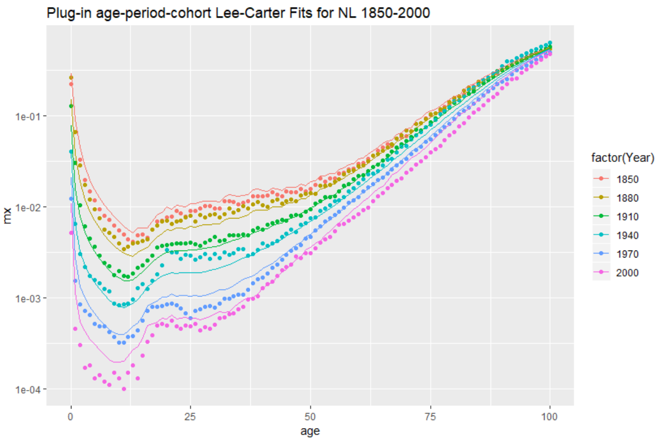


Table 10: MSE (w.r.t. mortality rates) of LC, plug-in age-period LC & plug-in age-period-cohort LC models.

	LC	Plug-in age-period LC	Plug-in age-period-cohort LC
MSE	5.153148e-08	3.250776e-06	2.837274e-06

6 Conclusion

Under the constraints $\sum_{x=0}^{X}\beta_x = 1, \sum_{t=1}^{T}\kappa_t = 0$, the Lee-Carter model is identifiable. The constraints $\sum_{t=1}^{T}\kappa_t = 0$ is automatically fulfilled during the estimation procedure, while the estimates need to be re-scaled to fulfill the constraint $\sum_{x=0}^{X}\beta_x = 1$. Moreover, the Lee-Carter estimation in the first step of the two-step Lee-Carter estimation procedure is justified by maximum likelihood estimation, if assume the errors are independent and identically distributed with distribution $N(0,1)$. However, imposing homoskedasticity and independence on the error structure is unrealistic, as the variance of the logarithm of the observed force of mortality is much greater at older ages than at younger ages, and the observed mortality rates are not uncorrelated. Furthermore, the estimation of the systematic non-random structure underpinning the Lee-Carter approach is based on the first set of singular vectors, this method has the potential for inflexibility with respect to age. In addition, the Lee-Carter model is fitted to the mortality rates rather than the actual total deaths, therefore, there is a lack of fit with respect to the actual total deaths, if κ is not re-estimated. On the other hand, The Lee-Carter age-period model with Poisson errors (PB predictor) doesnot impose homoskedasticity on the error structure of the Lee-Carter model and automatically ensures that total actual deaths and total expected deaths are identical for each year, hence avoids the need for κ to be re-estimated. Nevertheless, the estimates still need to be re-scaled to fulfill the constraints and the iterative process doesnot always converge. Another extension of the Lee-Carter model, the age-period-cohort Lee-Carter model, incorporates cohort effects, which could improve the performance of the Lee-Carter model. However, the imposed constraints for this model cannot insure the identifiability of this model. In terms of forecasting, when $\{\kappa_t\}$ follows from an AR(1) process and the sequence is non-stationary, the two step estimation can detect the true dynamics of κ_t's, if T is large enough. Despite this, using $\hat{\kappa}_t$ and **auto.arima** can fail to obtain the right model. Moreover, estimators of ARIMA models are conjectured to be inconsistent for the general Lee-Carter model when $\{\kappa_t\}$ follows from an ARIMA(p,d,q) model with $p + d + q > 1$.

Another issue with the Lee-Carter model is that the constraint $\sum_{t=1}^{T}\kappa_t = 0$ becomes a constraint on the possible realization of the stochastic process $\{\kappa_t\}$. The plug-in age-period Lee-Carter model and the plug-in age-period-cohort Lee-Carter model drop the constraint $\sum_{t=1}^{T}\kappa_t = 0$ and replace κ_t by time series models from the beginning. Moreover, these models are identifiable if $T \geq 2$ and $T \geq X + 2$ $(X > 0)$, respectively. However, estimates obtained using the objective function proposed by Beutner, Reese & Urbain (2017) and the R package **Rsolnp** do not fit the data. Besides, the estimation procedure is time-consuming if the data set is large. After adjusting the objective functions, the goodness of fit of the plug-in models is improved, but estimation procedure is still time-consuming. Moreover, the function **solnp** in the R package **Rsolnp** requires starting values and different starting values result in different estimates. One drawback of these plug-in models is that, the ARIMA process that $\{\kappa_t\}$ follows needs to be determined in advance, while the future mortality index κ_t is unobservable, therefore, the generated forecasts are questionable. Furthermore, the error terms e_t and ϵ_t in the plug-in age-period Lee-Carter model (e_{0t}, e_{1t} and ϵ_t in the plug-in age-period-cohort model) are assumed to be independent of each other, which could be extended such that they are correlated with each other.

This study examines the strength and weakness of the Lee-Carter model and some of its extensions, and offers useful insight into the performance of the plug-in Lee-Carter models. In terms of future work, the impact of different set of constraints for the Lee-Carter models, especially for the age-period-cohort Lee-Carter model, the impact of different R package or other program packages as well as the forecasting method for the plug-in age-period-cohort Lee-Carter model could be investigated.

Appendix

A. Tables

Table 11: Estimates of plug-in age-period model for NL 5x1 1850-2000.

$\hat{\alpha}_0$	-1.40099	$\hat{\alpha}_{12}$	-3.9743	$\hat{\beta}_0$	-411.782	$\hat{\beta}_{12}$	133.1086	$\hat{\mu}$	3.64E-05
$\hat{\alpha}_1$	-3.46428	$\hat{\alpha}_{13}$	-3.57452	$\hat{\beta}_1$	-348.672	$\hat{\beta}_{13}$	116.9352		
$\hat{\alpha}_2$	-5.03873	$\hat{\alpha}_{14}$	-3.14988	$\hat{\beta}_2$	-129.93	$\hat{\beta}_{14}$	106.6634		
$\hat{\alpha}_3$	-5.70788	$\hat{\alpha}_{15}$	-2.68118	$\hat{\beta}_3$	15.21864	$\hat{\beta}_{15}$	73.24938		
$\hat{\alpha}_4$	-5.48537	$\hat{\alpha}_{16}$	-2.23826	$\hat{\beta}_4$	74.92669	$\hat{\beta}_{16}$	57.05057		
$\hat{\alpha}_5$	-5.10786	$\hat{\alpha}_{17}$	-1.82098	$\hat{\beta}_5$	16.15687	$\hat{\beta}_{17}$	43.29408		
$\hat{\alpha}_6$	-4.99158	$\hat{\alpha}_{18}$	-1.41021	$\hat{\beta}_6$	-12.1705	$\hat{\beta}_{18}$	23.54538		
$\hat{\alpha}_7$	-4.8655	$\hat{\alpha}_{19}$	-1.06626	$\hat{\beta}_7$	-16.2189	$\hat{\beta}_{19}$	11.88247		
$\hat{\alpha}_8$	-4.72054	$\hat{\alpha}_{20}$	-0.77118	$\hat{\beta}_8$	1.062776	$\hat{\beta}_{20}$	2.151219		
$\hat{\alpha}_9$	-4.6162	$\hat{\alpha}_{21}$	-0.54142	$\hat{\beta}_9$	41.41313	$\hat{\beta}_{21}$	2.326648		
$\hat{\alpha}_{10}$	-4.50897	$\hat{\alpha}_{22}$	-0.36417	$\hat{\beta}_{10}$	96.91435	$\hat{\beta}_{22}$	-0.39778		
$\hat{\alpha}_{11}$	-4.25486	$\hat{\alpha}_{23}$	-0.24219	$\hat{\beta}_{11}$	116.2855	$\hat{\beta}_{23}$	-12.0143		

Table 12: Estimates of plug-in age-period model for NL 1x1 1850-2000.

$\hat{\alpha}_0$	0.367184	$\hat{\beta}_0$	0.017691	$\hat{\alpha}_{35}$	-3.01284	$\hat{\beta}_{35}$	0.01477	$\hat{\alpha}_{70}$	-2.22489	$\hat{\beta}_{70}$	0.004605
$\hat{\alpha}_1$	-0.25551	$\hat{\beta}_1$	0.023769	$\hat{\alpha}_{36}$	-3.01823	$\hat{\beta}_{36}$	0.014356	$\hat{\alpha}_{71}$	-2.26222	$\hat{\beta}_{71}$	0.00392
$\hat{\alpha}_2$	-0.92007	$\hat{\beta}_2$	0.02394	$\hat{\alpha}_{37}$	-3.01034	$\hat{\beta}_{37}$	0.014203	$\hat{\alpha}_{72}$	-2.00971	$\hat{\beta}_{72}$	0.00472
$\hat{\alpha}_3$	-1.55897	$\hat{\beta}_3$	0.022494	$\hat{\alpha}_{38}$	-2.83477	$\hat{\beta}_{38}$	0.014804	$\hat{\alpha}_{73}$	-1.94296	$\hat{\beta}_{73}$	0.00461
$\hat{\alpha}_4$	-1.96212	$\hat{\beta}_4$	0.021751	$\hat{\alpha}_{39}$	-2.88158	$\hat{\beta}_{39}$	0.014229	$\hat{\alpha}_{74}$	-1.85464	$\hat{\beta}_{74}$	0.004489
$\hat{\alpha}_5$	-2.26489	$\hat{\beta}_5$	0.021148	$\hat{\alpha}_{40}$	-2.96936	$\hat{\beta}_{40}$	0.013509	$\hat{\alpha}_{75}$	-1.79351	$\hat{\beta}_{75}$	0.004352
$\hat{\alpha}_6$	-2.49713	$\hat{\beta}_6$	0.020828	$\hat{\alpha}_{41}$	-3.05075	$\hat{\beta}_{41}$	0.012842	$\hat{\alpha}_{76}$	-1.79333	$\hat{\beta}_{76}$	0.003819
$\hat{\alpha}_7$	-2.71466	$\hat{\beta}_7$	0.020441	$\hat{\alpha}_{42}$	-2.97553	$\hat{\beta}_{42}$	0.012837	$\hat{\alpha}_{77}$	-1.71866	$\hat{\beta}_{77}$	0.00373
$\hat{\alpha}_8$	-2.87451	$\hat{\beta}_8$	0.020339	$\hat{\alpha}_{43}$	-3.09329	$\hat{\beta}_{43}$	0.01187	$\hat{\alpha}_{78}$	-1.67791	$\hat{\beta}_{78}$	0.003363
$\hat{\alpha}_9$	-3.05939	$\hat{\beta}_9$	0.019903	$\hat{\alpha}_{44}$	-3.18129	$\hat{\beta}_{44}$	0.011163	$\hat{\alpha}_{79}$	-1.61156	$\hat{\beta}_{79}$	0.003264
$\hat{\alpha}_{10}$	-3.22367	$\hat{\beta}_{10}$	0.019389	$\hat{\alpha}_{45}$	-3.08733	$\hat{\beta}_{45}$	0.011285	$\hat{\alpha}_{80}$	-1.52922	$\hat{\beta}_{80}$	0.003153
$\hat{\alpha}_{11}$	-3.40476	$\hat{\beta}_{11}$	0.018698	$\hat{\alpha}_{46}$	-3.17932	$\hat{\beta}_{46}$	0.010468	$\hat{\alpha}_{81}$	-1.57055	$\hat{\beta}_{81}$	0.002579
$\hat{\alpha}_{12}$	-3.59076	$\hat{\beta}_{12}$	0.017606	$\hat{\alpha}_{47}$	-3.24939	$\hat{\beta}_{47}$	0.009665	$\hat{\alpha}_{82}$	-1.36264	$\hat{\beta}_{82}$	0.003188
$\hat{\alpha}_{13}$	-3.77682	$\hat{\beta}_{13}$	0.016375	$\hat{\alpha}_{48}$	-3.1397	$\hat{\beta}_{48}$	0.009935	$\hat{\alpha}_{83}$	-1.27552	$\hat{\beta}_{83}$	0.003078
$\hat{\alpha}_{14}$	-3.64895	$\hat{\beta}_{14}$	0.016542	$\hat{\alpha}_{49}$	-3.27367	$\hat{\beta}_{49}$	0.008849	$\hat{\alpha}_{84}$	-1.25913	$\hat{\beta}_{84}$	0.002671
$\hat{\alpha}_{15}$	-3.74259	$\hat{\beta}_{15}$	0.015395	$\hat{\alpha}_{50}$	-3.1766	$\hat{\beta}_{50}$	0.008862	$\hat{\alpha}_{85}$	-1.19909	$\hat{\beta}_{85}$	0.002584
$\hat{\alpha}_{16}$	-3.9085	$\hat{\beta}_{16}$	0.013599	$\hat{\alpha}_{51}$	-3.18641	$\hat{\beta}_{51}$	0.008509	$\hat{\alpha}_{86}$	-1.08755	$\hat{\beta}_{86}$	0.002676
$\hat{\alpha}_{17}$	-3.67846	$\hat{\beta}_{17}$	0.014294	$\hat{\alpha}_{52}$	-3.04566	$\hat{\beta}_{52}$	0.008747	$\hat{\alpha}_{87}$	-1.09203	$\hat{\beta}_{87}$	0.002135
$\hat{\alpha}_{18}$	-3.47456	$\hat{\beta}_{18}$	0.014825	$\hat{\alpha}_{53}$	-3.09697	$\hat{\beta}_{53}$	0.0081	$\hat{\alpha}_{88}$	-1.04543	$\hat{\beta}_{88}$	0.001966
$\hat{\alpha}_{19}$	-3.41895	$\hat{\beta}_{19}$	0.014635	$\hat{\alpha}_{54}$	-3.07665	$\hat{\beta}_{54}$	0.007714	$\hat{\alpha}_{89}$	-0.97618	$\hat{\beta}_{89}$	0.001942
$\hat{\alpha}_{20}$	-3.22022	$\hat{\beta}_{20}$	0.015428	$\hat{\alpha}_{55}$	-3.15095	$\hat{\beta}_{55}$	0.006883	$\hat{\alpha}_{90}$	-0.93032	$\hat{\beta}_{90}$	0.001765
$\hat{\alpha}_{21}$	-3.35301	$\hat{\beta}_{21}$	0.01452	$\hat{\alpha}_{56}$	-3.02258	$\hat{\beta}_{56}$	0.007139	$\hat{\alpha}_{91}$	-0.89177	$\hat{\beta}_{91}$	0.001532
$\hat{\alpha}_{22}$	-3.15533	$\hat{\beta}_{22}$	0.015582	$\hat{\alpha}_{57}$	-3.02741	$\hat{\beta}_{57}$	0.006657	$\hat{\alpha}_{92}$	-0.84328	$\hat{\beta}_{92}$	0.001392
$\hat{\alpha}_{23}$	-3.18417	$\hat{\beta}_{23}$	0.015435	$\hat{\alpha}_{58}$	-2.95553	$\hat{\beta}_{58}$	0.006558	$\hat{\alpha}_{93}$	-0.79172	$\hat{\beta}_{93}$	0.001285
$\hat{\alpha}_{24}$	-3.23433	$\hat{\beta}_{24}$	0.015171	$\hat{\alpha}_{59}$	-2.97958	$\hat{\beta}_{59}$	0.006068	$\hat{\alpha}_{94}$	-0.75427	$\hat{\beta}_{94}$	0.0011
$\hat{\alpha}_{25}$	-3.08111	$\hat{\beta}_{25}$	0.015895	$\hat{\alpha}_{60}$	-2.85159	$\hat{\beta}_{60}$	0.006182	$\hat{\alpha}_{95}$	-0.69004	$\hat{\beta}_{95}$	0.001143
$\hat{\alpha}_{26}$	-3.11296	$\hat{\beta}_{26}$	0.015706	$\hat{\alpha}_{61}$	-2.86263	$\hat{\beta}_{61}$	0.005739	$\hat{\alpha}_{96}$	-0.67111	$\hat{\beta}_{96}$	0.000842
$\hat{\alpha}_{27}$	-3.03369	$\hat{\beta}_{27}$	0.016096	$\hat{\alpha}_{62}$	-2.754	$\hat{\beta}_{62}$	0.005817	$\hat{\alpha}_{97}$	-0.63029	$\hat{\beta}_{97}$	0.000767
$\hat{\alpha}_{28}$	-3.01541	$\hat{\beta}_{28}$	0.016088	$\hat{\alpha}_{63}$	-2.67062	$\hat{\beta}_{63}$	0.005764	$\hat{\alpha}_{98}$	-0.59254	$\hat{\beta}_{98}$	0.000655
$\hat{\alpha}_{29}$	-3.03848	$\hat{\beta}_{29}$	0.01584	$\hat{\alpha}_{64}$	-2.60891	$\hat{\beta}_{64}$	0.005548	$\hat{\alpha}_{99}$	-0.54965	$\hat{\beta}_{99}$	0.0006
$\hat{\alpha}_{30}$	-3.06843	$\hat{\beta}_{30}$	0.015496	$\hat{\alpha}_{65}$	-2.48555	$\hat{\beta}_{65}$	0.005737	$\hat{\alpha}_{100}$	-0.52865	$\hat{\beta}_{100}$	0.000418
$\hat{\alpha}_{31}$	-3.13614	$\hat{\beta}_{31}$	0.014966	$\hat{\alpha}_{66}$	-2.49799	$\hat{\beta}_{66}$	0.005066	$\hat{\mu}$	-1.23271		
$\hat{\alpha}_{32}$	-2.92801	$\hat{\beta}_{32}$	0.015797	$\hat{\alpha}_{67}$	-2.50788	$\hat{\beta}_{67}$	0.004644	$\hat{c}$	-89.1495		
$\hat{\alpha}_{33}$	-2.86437	$\hat{\beta}_{33}$	0.01593	$\hat{\alpha}_{68}$	-2.29487	$\hat{\beta}_{68}$	0.005326	$\hat{\sigma}^2_e$	1.176203		
$\hat{\alpha}_{34}$	-2.92726	$\hat{\beta}_{34}$	0.015406	$\hat{\alpha}_{69}$	-2.30087	$\hat{\beta}_{69}$	0.004793	$\hat{\sigma}^2_\epsilon$	0.037528		

Table 13: Estimates of plug-in age-period-cohort model for NL 1x1 1850-2000.

$\hat{\alpha}_0$	-0.90962	$\hat{\beta}_0^{(1)}$	0.090899	$\hat{\beta}_0^{(0)}$	-0.00305	$\hat{\alpha}_{52}$	-1.89299831	$\hat{\beta}_{52}^{(1)}$	-0.08164	$\hat{\beta}_{52}^{(0)}$	0.03462
$\hat{\alpha}_1$	-1.1729	$\hat{\beta}_1^{(1)}$	0.092274	$\hat{\beta}_1^{(0)}$	0.004166	$\hat{\alpha}_{53}$	-1.99754247	$\hat{\beta}_{53}^{(1)}$	-0.08099	$\hat{\beta}_{53}^{(0)}$	0.033513
$\hat{\alpha}_2$	-1.25635	$\hat{\beta}_2^{(1)}$	0.071163	$\hat{\beta}_2^{(0)}$	0.010414	$\hat{\alpha}_{54}$	-1.83197747	$\hat{\beta}_{54}^{(1)}$	-0.09387	$\hat{\beta}_{54}^{(0)}$	0.036831
$\hat{\alpha}_3$	-1.39368	$\hat{\beta}_3^{(1)}$	0.04888	$\hat{\beta}_3^{(0)}$	0.014974	$\hat{\alpha}_{55}$	-1.75923201	$\hat{\beta}_{55}^{(1)}$	-0.10695	$\hat{\beta}_{55}^{(0)}$	0.039719
$\hat{\alpha}_4$	-1.10389	$\hat{\beta}_4^{(1)}$	0.020498	$\hat{\beta}_4^{(0)}$	0.022125	$\hat{\alpha}_{56}$	-1.82552715	$\hat{\beta}_{56}^{(1)}$	-0.09518	$\hat{\beta}_{56}^{(0)}$	0.036546
$\hat{\alpha}_5$	-1.41079	$\hat{\beta}_5^{(1)}$	0.018359	$\hat{\beta}_5^{(0)}$	0.021957	$\hat{\alpha}_{57}$	-1.72064395	$\hat{\beta}_{57}^{(1)}$	-0.1068	$\hat{\beta}_{57}^{(0)}$	0.0393
$\hat{\alpha}_6$	-1.39146	$\hat{\beta}_6^{(1)}$	0.007075	$\hat{\beta}_6^{(0)}$	0.02476	$\hat{\alpha}_{58}$	-1.9532168	$\hat{\beta}_{58}^{(1)}$	-0.08427	$\hat{\beta}_{58}^{(0)}$	0.032735
$\hat{\alpha}_7$	-1.34895	$\hat{\beta}_7^{(1)}$	-0.00506	$\hat{\beta}_7^{(0)}$	0.027687	$\hat{\alpha}_{59}$	-1.97475058	$\hat{\beta}_{59}^{(1)}$	-0.08781	$\hat{\beta}_{59}^{(0)}$	0.032999
$\hat{\alpha}_8$	-1.354	$\hat{\beta}_8^{(1)}$	-0.01258	$\hat{\beta}_8^{(0)}$	0.029673	$\hat{\alpha}_{60}$	-2.24883758	$\hat{\beta}_{60}^{(1)}$	-0.05422	$\hat{\beta}_{60}^{(0)}$	0.023664
$\hat{\alpha}_9$	-1.18926	$\hat{\beta}_9^{(1)}$	-0.02894	$\hat{\beta}_9^{(0)}$	0.033712	$\hat{\alpha}_{61}$	-2.32126075	$\hat{\beta}_{61}^{(1)}$	-0.05138	$\hat{\beta}_{61}^{(0)}$	0.022162
$\hat{\alpha}_{10}$	-1.48104	$\hat{\beta}_{10}^{(1)}$	-0.02647	$\hat{\beta}_{10}^{(0)}$	0.032315	$\hat{\alpha}_{62}$	-2.32282223	$\hat{\beta}_{62}^{(1)}$	-0.04303	$\hat{\beta}_{62}^{(0)}$	0.019893
$\hat{\alpha}_{11}$	-1.59785	$\hat{\beta}_{11}^{(1)}$	-0.03235	$\hat{\beta}_{11}^{(0)}$	0.033018	$\hat{\alpha}_{63}$	-2.40717809	$\hat{\beta}_{63}^{(1)}$	-0.0296	$\hat{\beta}_{63}^{(0)}$	0.015874
$\hat{\alpha}_{12}$	-1.47961	$\hat{\beta}_{12}^{(1)}$	-0.04896	$\hat{\beta}_{12}^{(0)}$	0.036275	$\hat{\alpha}_{64}$	-2.12849574	$\hat{\beta}_{64}^{(1)}$	-0.05103	$\hat{\beta}_{64}^{(0)}$	0.021869
$\hat{\alpha}_{13}$	-1.61191	$\hat{\beta}_{13}^{(1)}$	-0.05589	$\hat{\beta}_{13}^{(0)}$	0.036568	$\hat{\alpha}_{65}$	-2.29260075	$\hat{\beta}_{65}^{(1)}$	-0.02448	$\hat{\beta}_{65}^{(0)}$	0.014514
$\hat{\alpha}_{14}$	-1.30546	$\hat{\beta}_{14}^{(1)}$	-0.06465	$\hat{\beta}_{14}^{(0)}$	0.03925	$\hat{\alpha}_{66}$	-2.04736119	$\hat{\beta}_{66}^{(1)}$	-0.04944	$\hat{\beta}_{66}^{(0)}$	0.021076
$\hat{\alpha}_{15}$	-1.38885	$\hat{\beta}_{15}^{(1)}$	-0.06883	$\hat{\beta}_{15}^{(0)}$	0.039025	$\hat{\alpha}_{67}$	-2.15558651	$\hat{\beta}_{67}^{(1)}$	-0.0435	$\hat{\beta}_{67}^{(0)}$	0.018478
$\hat{\alpha}_{16}$	-1.52922	$\hat{\beta}_{16}^{(1)}$	-0.07531	$\hat{\beta}_{16}^{(0)}$	0.038575	$\hat{\alpha}_{68}$	-2.21077598	$\hat{\beta}_{68}^{(1)}$	-0.01864	$\hat{\beta}_{68}^{(0)}$	0.012053
$\hat{\alpha}_{17}$	-1.50424	$\hat{\beta}_{17}^{(1)}$	-0.06627	$\hat{\beta}_{17}^{(0)}$	0.036938	$\hat{\alpha}_{69}$	-1.98748443	$\hat{\beta}_{69}^{(1)}$	-0.0445	$\hat{\beta}_{69}^{(0)}$	0.018806
$\hat{\alpha}_{18}$	-1.52448	$\hat{\beta}_{18}^{(1)}$	-0.05628	$\hat{\beta}_{18}^{(0)}$	0.034879	$\hat{\alpha}_{70}$	-2.23191489	$\hat{\beta}_{70}^{(1)}$	-0.00728	$\hat{\beta}_{70}^{(0)}$	0.008099
$\hat{\alpha}_{19}$	-1.53704	$\hat{\beta}_{19}^{(1)}$	-0.0546	$\hat{\beta}_{19}^{(0)}$	0.034249	$\hat{\alpha}_{71}$	-1.99067568	$\hat{\beta}_{71}^{(1)}$	-0.03946	$\hat{\beta}_{71}^{(0)}$	0.016528
$\hat{\alpha}_{20}$	-1.48533	$\hat{\beta}_{20}^{(1)}$	-0.0479	$\hat{\beta}_{20}^{(0)}$	0.033346	$\hat{\alpha}_{72}$	-2.33074814	$\hat{\beta}_{72}^{(1)}$	0.030163	$\hat{\beta}_{72}^{(0)}$	-0.00242
$\hat{\alpha}_{21}$	-1.34425	$\hat{\beta}_{21}^{(1)}$	-0.06292	$\hat{\beta}_{21}^{(0)}$	0.036594	$\hat{\alpha}_{73}$	-2.17421185	$\hat{\beta}_{73}^{(1)}$	0.017672	$\hat{\beta}_{73}^{(0)}$	0.000861
$\hat{\alpha}_{22}$	-1.3573	$\hat{\beta}_{22}^{(1)}$	-0.05301	$\hat{\beta}_{22}^{(0)}$	0.035083	$\hat{\alpha}_{74}$	-2.37792635	$\hat{\beta}_{74}^{(1)}$	0.059565	$\hat{\beta}_{74}^{(0)}$	-0.01102
$\hat{\alpha}_{23}$	-1.37093	$\hat{\beta}_{23}^{(1)}$	-0.05595	$\hat{\beta}_{23}^{(0)}$	0.035667	$\hat{\alpha}_{75}$	-2.29246296	$\hat{\beta}_{75}^{(1)}$	0.055727	$\hat{\beta}_{75}^{(0)}$	-0.01032
$\hat{\alpha}_{24}$	-1.39468	$\hat{\beta}_{24}^{(1)}$	-0.0592	$\hat{\beta}_{24}^{(0)}$	0.036278	$\hat{\alpha}_{76}$	-2.33391195	$\hat{\beta}_{76}^{(1)}$	0.066763	$\hat{\beta}_{76}^{(0)}$	-0.014
$\hat{\alpha}_{25}$	-1.49896	$\hat{\beta}_{25}^{(1)}$	-0.04695	$\hat{\beta}_{25}^{(0)}$	0.033747	$\hat{\alpha}_{77}$	-2.30857682	$\hat{\beta}_{77}^{(1)}$	0.075671	$\hat{\beta}_{77}^{(0)}$	-0.01675
$\hat{\alpha}_{26}$	-1.3983	$\hat{\beta}_{26}^{(1)}$	-0.05513	$\hat{\beta}_{26}^{(0)}$	0.035878	$\hat{\alpha}_{78}$	-2.44236877	$\hat{\beta}_{78}^{(1)}$	0.110061	$\hat{\beta}_{78}^{(0)}$	-0.02675
$\hat{\alpha}_{27}$	-1.29417	$\hat{\beta}_{27}^{(1)}$	-0.05746	$\hat{\beta}_{27}^{(0)}$	0.037021	$\hat{\alpha}_{79}$	-2.4237646	$\hat{\beta}_{79}^{(1)}$	0.121256	$\hat{\beta}_{79}^{(0)}$	-0.03022
$\hat{\alpha}_{28}$	-1.20619	$\hat{\beta}_{28}^{(1)}$	-0.06296	$\hat{\beta}_{28}^{(0)}$	0.038537	$\hat{\alpha}_{80}$	-2.20766095	$\hat{\beta}_{80}^{(1)}$	0.104547	$\hat{\beta}_{80}^{(0)}$	-0.02552
$\hat{\alpha}_{29}$	-1.30673	$\hat{\beta}_{29}^{(1)}$	-0.06164	$\hat{\beta}_{29}^{(0)}$	0.037794	$\hat{\alpha}_{81}$	-2.07445313	$\hat{\beta}_{81}^{(1)}$	0.077968	$\hat{\beta}_{81}^{(0)}$	-0.01889
$\hat{\alpha}_{30}$	-1.34072	$\hat{\beta}_{30}^{(1)}$	-0.06334	$\hat{\beta}_{30}^{(0)}$	0.037927	$\hat{\alpha}_{82}$	-1.65239804	$\hat{\beta}_{82}^{(1)}$	0.034286	$\hat{\beta}_{82}^{(0)}$	-0.00583
$\hat{\alpha}_{31}$	-1.42692	$\hat{\beta}_{31}^{(1)}$	-0.06507	$\hat{\beta}_{31}^{(0)}$	0.03775	$\hat{\alpha}_{83}$	-2.08418645	$\hat{\beta}_{83}^{(1)}$	0.144539	$\hat{\beta}_{83}^{(0)}$	-0.03716
$\hat{\alpha}_{32}$	-1.32457	$\hat{\beta}_{32}^{(1)}$	-0.05945	$\hat{\beta}_{32}^{(0)}$	0.037316	$\hat{\alpha}_{84}$	-2.03706617	$\hat{\beta}_{84}^{(1)}$	0.148055	$\hat{\beta}_{84}^{(0)}$	-0.0387
$\hat{\alpha}_{33}$	-1.2334	$\hat{\beta}_{33}^{(1)}$	-0.06267	$\hat{\beta}_{33}^{(0)}$	0.0384	$\hat{\alpha}_{85}$	-1.3891222	$\hat{\beta}_{85}^{(1)}$	0.022811	$\hat{\beta}_{85}^{(0)}$	-0.00318
$\hat{\alpha}_{34}$	-1.27658	$\hat{\beta}_{34}^{(1)}$	-0.06713	$\hat{\beta}_{34}^{(0)}$	0.038899	$\hat{\alpha}_{86}$	-1.2525586	$\hat{\beta}_{86}^{(1)}$	0.015266	$\hat{\beta}_{86}^{(0)}$	-0.00098
$\hat{\alpha}_{35}$	-1.47054	$\hat{\beta}_{35}^{(1)}$	-0.06512	$\hat{\beta}_{35}^{(0)}$	0.037341	$\hat{\alpha}_{87}$	-1.67135435	$\hat{\beta}_{87}^{(1)}$	0.132285	$\hat{\beta}_{87}^{(0)}$	-0.0348
$\hat{\alpha}_{36}$	-1.51136	$\hat{\beta}_{36}^{(1)}$	-0.06444	$\hat{\beta}_{36}^{(0)}$	0.036825	$\hat{\alpha}_{88}$	-1.51347303	$\hat{\beta}_{88}^{(1)}$	0.10955	$\hat{\beta}_{88}^{(0)}$	-0.02867
$\hat{\alpha}_{37}$	-1.53132	$\hat{\beta}_{37}^{(1)}$	-0.06547	$\hat{\beta}_{37}^{(0)}$	0.036862	$\hat{\alpha}_{89}$	-2.07738514	$\hat{\beta}_{89}^{(1)}$	0.306885	$\hat{\beta}_{89}^{(0)}$	-0.08507
$\hat{\alpha}_{38}$	-1.49323	$\hat{\beta}_{38}^{(1)}$	-0.05831	$\hat{\beta}_{38}^{(0)}$	0.035665	$\hat{\alpha}_{90}$	-1.67071383	$\hat{\beta}_{90}^{(1)}$	0.218425	$\hat{\beta}_{90}^{(0)}$	-0.06007
$\hat{\alpha}_{39}$	-1.47311	$\hat{\beta}_{39}^{(1)}$	-0.06497	$\hat{\beta}_{39}^{(0)}$	0.036819	$\hat{\alpha}_{91}$	-1.6140858	$\hat{\beta}_{91}^{(1)}$	0.238375	$\hat{\beta}_{91}^{(0)}$	-0.06598
$\hat{\alpha}_{40}$	-1.53362	$\hat{\beta}_{40}^{(1)}$	-0.07053	$\hat{\beta}_{40}^{(0)}$	0.037331	$\hat{\alpha}_{92}$	-1.38605292	$\hat{\beta}_{92}^{(1)}$	0.192693	$\hat{\beta}_{92}^{(0)}$	-0.0532
$\hat{\alpha}_{41}$	-1.5699	$\hat{\beta}_{41}^{(1)}$	-0.07526	$\hat{\beta}_{41}^{(0)}$	0.037975	$\hat{\alpha}_{93}$	-1.14888391	$\hat{\beta}_{93}^{(1)}$	0.132334	$\hat{\beta}_{93}^{(0)}$	-0.03616
$\hat{\alpha}_{42}$	-1.44329	$\hat{\beta}_{42}^{(1)}$	-0.0804	$\hat{\beta}_{42}^{(0)}$	0.039481	$\hat{\alpha}_{94}$	-1.01770362	$\hat{\beta}_{94}^{(1)}$	0.108327	$\hat{\beta}_{94}^{(0)}$	-0.02949
$\hat{\alpha}_{43}$	-1.49668	$\hat{\beta}_{43}^{(1)}$	-0.08734	$\hat{\beta}_{43}^{(0)}$	0.040303	$\hat{\alpha}_{95}$	-1.15644121	$\hat{\beta}_{95}^{(1)}$	0.237776	$\hat{\beta}_{95}^{(0)}$	-0.06655
$\hat{\alpha}_{44}$	-1.40527	$\hat{\beta}_{44}^{(1)}$	-0.1032	$\hat{\beta}_{44}^{(0)}$	0.043718	$\hat{\alpha}_{96}$	-0.94067626	$\hat{\beta}_{96}^{(1)}$	0.161722	$\hat{\beta}_{96}^{(0)}$	-0.04504
$\hat{\alpha}_{45}$	-1.54623	$\hat{\beta}_{45}^{(1)}$	-0.09047	$\hat{\beta}_{45}^{(0)}$	0.04025	$\hat{\alpha}_{97}$	-1.00645691	$\hat{\beta}_{97}^{(1)}$	0.29898	$\hat{\beta}_{97}^{(0)}$	-0.08431
$\hat{\alpha}_{46}$	-1.67718	$\hat{\beta}_{46}^{(1)}$	-0.09093	$\hat{\beta}_{46}^{(0)}$	0.039364	$\hat{\alpha}_{98}$	-0.98510563	$\hat{\beta}_{98}^{(1)}$	0.420031	$\hat{\beta}_{98}^{(0)}$	-0.11896
$\hat{\alpha}_{47}$	-1.72768	$\hat{\beta}_{47}^{(1)}$	-0.09404	$\hat{\beta}_{47}^{(0)}$	0.039442	$\hat{\alpha}_{99}$	-0.84774901	$\hat{\beta}_{99}^{(1)}$	0.450686	$\hat{\beta}_{99}^{(0)}$	-0.12783
$\hat{\alpha}_{48}$	-1.96021	$\hat{\beta}_{48}^{(1)}$	-0.07426	$\hat{\beta}_{48}^{(0)}$	0.033933	$\hat{\alpha}_{100}$	-0.71486004	$\hat{\beta}_{100}^{(1)}$	0.547284	$\hat{\beta}_{100}^{(0)}$	-0.15548
$\hat{\alpha}_{49}$	-1.96415	$\hat{\beta}_{49}^{(1)}$	-0.08685	$\hat{\beta}_{49}^{(0)}$	0.036057	$\hat{\mu 1}$	-0.273578	$\hat{\mu 0}$	-0.9599	σ_ϵ^2	0.01385
$\hat{\alpha}_{50}$	-2.01789	$\hat{\beta}_{50}^{(1)}$	-0.0772	$\hat{\beta}_{50}^{(0)}$	0.033526	$\hat{c}1$	-7.486447	$\hat{c}0$	-27.307		
$\hat{\alpha}_{51}$	-2.1884	$\hat{\beta}_{51}^{(1)}$	-0.06905	$\hat{\beta}_{51}^{(0)}$	0.030594	$\hat{\sigma}_{e1}^2$	6.08E-07	$\hat{\sigma}_{e0}^2$	0.16059		

B. R codes

```
rm(list = ls())
library(Rsolnp)
library(dplyr)
library(stringr)
library(ggplot2)
library(forecast)
library(dlm)

data_rates <- read.table('nl.txt',header=T)
data_rates <- mutate(data_rates,age=as.numeric(str_extract(data_rates$Age, "[0-9]+")))
mydata_rates <- select(data_rates,Year,age,mx)%>%filter(age<=100)
rates <- spread(mydata_rates,Year,mx)[,-1]
data_pop <- read.table('nl_pop.txt',header=T)
data_pop <- mutate(data_pop,age=as.numeric(str_extract(data_pop$Age, "[0-9]+")))
mydata_pop<- select(data_pop,Year,age,Total)%>%filter(age<=100)
pop <- spread(mydata_pop,Year,Total)[,-1]
ages <- unique(mydata_rates$age)
years <- unique(mydata_rates$Year)
K <- length(years)
data_deaths <- read.table('nl_deaths.txt',header=T)%>%mutate(age=as.numeric(str_extract(data_rates$Age, "[0-9]+")))
data_deaths <- data_deaths %>% filter(age<=100)
mydata <- mutate(mydata_rates,pop=mydata_pop$Total,deaths=data_deaths$Total)
mydata_s <- filter(mydata,Year<=2000)
mydata_f<- filter(mydata,Year>2000)

rates <- mydata_s$mx
t <- unique(mydata_s$Year)
x <- unique(mydata_s$age)
X <- length(x)
T <- length(t)

ln_data <- matrix(log(rates), X, T)
rownames(ln_data) <- x
colnames(ln_data) <- t

#############################
#######
#plug in age-period LC with adjusted objective function#
#######
minimi <- function(theta,X=101,T=151){
  alpha <- theta[1:X]
  beta <- theta[(X+1):(2*X)]
  mu<- theta[2*X+1]
  c <- theta[2*X+2]
  sum1=0

  for(x in 1:X){
    for(t in 1:T){
      sum1 <- sum1 +1/t*(ln_data[x,t]-(alpha[x]+beta[x]*c+beta[x]*mu*t))^2

    }
  }
  return(sum1)
}

###############################
equal <- function(theta){
  sum(theta[(X+1):(2*X)])
}
```

```
lb <- rep(-1000,2*X+2)
ub <- rep(10000,2*X+2)
##
estimates<- solnp(rep(1/X,2*X+2),minimi,eqfun=equal,eqB=1,LB=lb,UB=ub)
pars <- estimates$pars
##########
minimi2 <- function(theta2,X=101,T=151){
  alpha <- pars[1:X]
  beta <- pars[(X+1):(2*X)]
  mu <- pars[2*X+1]
  c <- pars[2*X+2]
  sigma_eSQ <- theta2[1]
  sigma_epsiSQ <- theta2[2]
  sum2=0

  for(x in 1:X){
    for(t in 1:T){
      for(y in 1:X){
        for(s in 1:T){
          sum2=sum2+1/min(s,t)*(ln_data[x,t]*ln_data[y,s]-
                                -beta[x]*beta[y]*sigma_eSQ*min(s,t)-(x==y&t==s)*sigma_epsiSQ-
                                 (alpha[x]+beta[x]*c+beta[x]*mu*t)*(alpha[y]+beta[y]*c+beta[y]*mu*s))^2
        }
      }
    }
  }
  return(sum2)
}

lb2 <- rep(0,2)
ub2 <- rep(10000,2)
##
estimates2<- solnp(rep(1/X,2),minimi2,LB=lb2,UB=ub2)
 pars2 <- estimates2$pars

################################
a <- pars[1:X]
b <- pars[(X+1):(2*X)]
mu <- pars[2*X+1]
c <- pars[2*X+2]
sigma_eSQ <- pars2[1]
sigma_epsiSQ <- pars2[2]
################################
fit2 <- NaN
for (t in 1:T){
  for (x in 1:X){
    fit2<- c(fit2,exp(a[x]+b[x]*c+b[x]*mu*t))
  }
}
fit2 <- fit2[-1]

datafit<-data.frame(mydata_s,fit2=fit2)

data_plot <- filter(datafit,Year == 1850| Year == 1880| Year == 1910| Year == 1940| Year == 1970| Year == 2000)

ggplot(data_plot, aes(age, mx,color=factor(Year))) + geom_point() +
  geom_line(aes(age,fit2,color=factor(Year))) + scale_y_log10() +
  ggtitle("Plug-in age-period Lee-Carter Fits for NL 1850-2000")

########################
#forecasting#
Zeta <- matrix(NaN,2,152)
```

```
Zeta[,1]<- c(mu+c,1)
Psi<- matrix(c((mu+c)^2,mu+c,mu+c,1),nrow=2,byrow=T)
Pi <- matrix(c((mu+c)^2+sigma_eSQ,mu+c,mu+c,1),nrow=2,byrow=T)
Phi <- matrix(c(sigma_eSQ,0,0,0),nrow=2,byrow=T)
R <- diag(rep(sigma_epsiSQ,101),nrow=101)
Q <- matrix(c(sigma_eSQ,0,0,0),nrow=2,byrow=T)
G <- cbind(b,a)
F <- matrix(c(1,mu+c,0,1),nrow=2,byrow=TRUE)
for(i in 1:151){
  delta <- G%*%Phi%*%t(G)+R
  Xi <- F%*%Phi%*%t(G)
  temp_Z <- F%*%Zeta[,i]+Xi%*%solve(delta)%*%(ln_data[,1]-G%*%Zeta[,i])
  Zeta[,i+1] <- temp_Z
  temp_Pi<- F%*%Pi%*%t(F)+Q*(i-1)
  Pi <- temp_Pi
  temp_Psi <- F%*%Psi%*%t(F)+Xi%*%solve(delta)%*%t(Xi)
  Psi <- temp_Psi
  Phi <- Pi-Psi
}

Zeta_2001 <-Zeta[,152]
forecasts <- exp(G%*%Zeta_2001)
for(i in 2:16){
  Fh <- matrix(c(1,(i-1)*mu+c,0,1),nrow=2,byrow=T)
  temp <- exp(G%*%Fh%*%Zeta_2001)
  forecasts <- cbind(forecasts,temp)
}

later <- NaN
for(i in 1:16){
  later<- c(later,forecasts[,i])
}
later <- later[-1]

#year 2001#
years <- mutate(mydata_f,forecast=later)%>%filter(Year==2001)
ggplot(years, aes(age, mx)) + scale_y_log10() +
  geom_point() + geom_line(aes(age, forecast)) +
  ggtitle("Observed and Forecast for the year 2001")

#year 2002#
years <- mutate(mydata_f,forecast=later)%>%filter(Year==2002)
ggplot(years, aes(age, mx)) + scale_y_log10() +
  geom_point() + geom_line(aes(age, forecast)) +
  ggtitle("Observed and Forecast for the year 2002")

#year 2003#
years <- mutate(mydata_f,forecast=later)%>%filter(Year==2003)
ggplot(years, aes(age, mx)) + scale_y_log10() +
  geom_point() + geom_line(aes(age, forecast)) +
  ggtitle("Observed and Forecast for the year 2003")

#year 2016#
years <- mutate(mydata_f,forecast=later)%>%filter(Year==2016)
ggplot(years, aes(age, mx)) + scale_y_log10() +
  geom_point() + geom_line(aes(age, forecast)) +
  ggtitle("Observed and Forecast for the year 2016")
```

References

Beutner, E., 2019. *Life Insurance II (lecture slides 2019)*. Maastricht University, Maastricht.

Beutner, E., Reese, S., & Urbain J., 2016. *Identifiability issues of age-period and age-period-cohort models of the Lee-Carter type*. Insurance: Mathematics and Economics, 75, 117-125.

Booth,H., Maindonald, J. & Smith, L., 2002. *Applying Lee-Carter under conditions of variable mortality decline*. Population Studies, 56:3,325-336.

Brockwell, P.J., & Davis, R.A, 1897. *Time Series: Theory and Methods*, first ed.. Springer-Verlag, New York.

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

Butt, Z., Haberman, S. & Shang, H.L., 2015. *The ilc Package*. Retrieved on 27 July 2019 from `https://www.researchgate.net/publication/271013835_The_ilc_Package_in_R_Generalised_Lee-Carter_models_using_iterative_fitting_algorithms`.

Lee, R.D., Carter, L.R., 1992. *Modeling and foreasting U.S. mortality.* Journal of American statistic association, 87, 659-671.

Leng, X., Peng, L., 2016. *Inference pitfalls in Lee-Carter model for forecasting mortality.* Insurance: Mathematics and Economics, 70, 58-65.

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

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

Renshaw, A.E., Haberman, S., 2008. *Mortality, longevity and experiments with the Lee-Carter model*. Lifetime Data Anal (2008) 14:286-315, DOI: 10.1007/s10985-008-9084-2.

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