Available online at www.sciencedirect.com
ScienceDirect
Insurance: Mathematics and Economics 42 (2008) 492–504
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/ime

Estimating the term structure of mortality[☆]

Norbert Hári[a], Anja De Waegenaere[b,*], Bertrand Melenberg[b,a], Theo E. Nijman[a]

[a] Finance Department, Tilburg University, CentER for Economic Research and Netspar, The Netherlands
[b] Department of Econometrics and OR, Tilburg University, CentER for Economic Research and Netspar, The Netherlands

Received September 2006; received in revised form January 2007; accepted 7 January 2007

Abstract

In modeling and forecasting mortality the Lee–Carter approach is the benchmark methodology. In many empirical applications the Lee–Carter approach results in a model that describes the log central death rates by means of linear trends. However, due to the volatility in (past) mortality data, the estimation of these trends, and, thus, the forecasts based on them, might be rather sensitive to the sample period employed. We allow for time-varying trends, depending on a few underlying factors, to make the estimates of the future trends less sensitive to the sampling period. We formulate our model in a state-space framework, and use the Kalman filtering technique to estimate it. We illustrate our model using Dutch mortality data.

JEL classification: C15; C51; C52; C53; J11

Keywords: Mortality forecasting; State-space modeling; Parameter risk

1. Introduction

For life-related insurance products, one can distinguish two types of actuarial risk. First, institutions offering products depending on the lifetime of an individual face risk, simply because lifetime is uncertain. However, it is well known that this type of risk reduces significantly when the portfolio size is increased. Secondly, mortality patterns may change over time due to, for example, improvements in the standards of living and lifestyle or better prospects in the medical system. This source of risk clearly cannot be diversified away by increasing the portfolio size. As a consequence, changes in survival probabilities can have a major effect on, for example, fair premiums for life insurance or funding ratios for pension funds. Therefore, forecasting future mortality risk is in the interest of insurance companies and pension funds.

Several methods for capturing the behavior of mortality rates over time and for forecasting future mortality have been developed. The literature evolved along several directions. The *deterministic trend* approach fits curves as a function of age and time to approximate mortality rates. Fitting curves to mortality rates goes back to Gompertz or Makeham in the 19th century. These early efforts tried to fit part of the mortality curve by considering only the age dimension, typically the middle and elderly aged. Heligman and Pollard (1980) already fitted a curve to the entire age range, but they did not estimate the time effect either. Most recent models fit curves to mortality rates in both the age and time dimension. For instance, Renshaw et al. (1996) use polynomials to describe the age and time evolution of mortality changes. These models give a very accurate in-sample fit. However, a main disadvantage of this deterministic trend approach is that the accurate in-sample fit is translated into quite small prediction intervals, when extrapolated out of sample, but such accurate predictions do not seem to be very realistic, also because of the model uncertainty that is usually not taken into account.

The *stochastic trend* methodology seems to be a more parsimonious approach, which tries to explain the variability of mortality rates with a low number of unobserved latent

[☆] We thank Frank de Jong, Bas Werker, an anonymous referee, the seminar participants of the 10th International Congress on Insurance: Mathematics and Economics, and of Netspar Pension Days for useful comments and suggestions.
[*] Corresponding address: Department of Econometrics and OR, Tilburg University, CentER for Economic Research and Netspar, PO Box 90153, 5000LE Tilburg, The Netherlands. Tel.: +31 13 4662913; fax: +31 13 4663280.
E-mail address: A.M.B.DeWaegenaere@uvt.nl (A. De Waegenaere).

0167-6687/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2007.01.011

factors: death rates are explained as a function of time-varying unobserved state variables and age-specific parameters, which describe the relative sensitivities of individual age groups to the change in the underlying unobserved state variables. The stochastic trend approach was first introduced for mortality forecasts in Lee and Carter (1992). They explore the time series behavior of mortality movements between age groups by using a single latent factor which is responsible for describing the general level of log mortality. Log central death rates are modeled as the sum of a time invariant, age-specific constant, and the product of an age-specific time invariant component and the time-varying latent factor. The age-specific component represents the sensitivity of an individual age group to the general level of mortality changes. The estimation of the model proceeds in several steps. First, singular value decomposition (SVD) is used to retrieve the underlying factor. Second, the age-specific parameters are estimated by means of ordinary least squares. Then the latent factor is re-estimated while keeping age-specific parameters from the first step constant, in order to guarantee that the sum of the implied number of deaths equals the sum of the actual number of deaths in each time period. Finally, ARIMA modeling is used to fit a time series process to the latent variable, which can be used to make forecasts. In case of Lee and Carter (1992) the time process of the latent factor turned out to be a random walk with drift, implying that its forecast is just a linear trend, but with a prediction interval much wider than obtained in case of a deterministic trend approach.

A whole strand of literature evolved from the original Lee–Carter approach, see, for example, Lee and Miller (2001), Carter and Prskawetz (2001), Booth et al. (2002), Brouhns et al. (2002), and Renshaw and Haberman (2003a,b) to mention just a few. Recently, Girosi and King (2005) proposed a reformulation of the empirically quite often found version of the Lee and Carter (1992)-model, which is the version having a single latent factor, following a random walk with drift. In this reformulation the log central death rates (or some other way to measure log mortalities) are directly modeled as random walks with drift, making estimation and forecasting a rather straightforward exercise in econometrics, simplifying considerably the original Lee and Carter (1992)-estimation and prediction approach.

We take this reformulation by Girosi and King (2005) as our starting point. When using actual mortality data to estimate this version of the Lee and Carter (1992)-model, we make the following observation. First, the typical sample period is rather short, usually starting somewhere in the nineteenth century, resulting in only around 150 annual observations (or even less). Secondly, the observed mortality data turns out to be quite volatile, particularly, during the nineteenth century, but also around, for instance, the first and second world war. This implies that the estimation of the drift term in the Girosi and King (2005) reformulation – the slope of the long run trend – might be rather sensitive to the sample period used in estimation, making also the long run forecasts sensitive to the sample period.

To account for this sensitivity, we propose to extend the Girosi and King (2005) formulation of the Lee and Carter (1992)-model by making the drift term time dependent. We postulate that this time dependent drift term is a (time independent) affine transformation of a few underlying (time-varying) latent factors, which capture the time movements, common to all different age groups. The underlying latent factors are assumed to have a long run zero mean, but their short run sample means might deviate from zero. These non-zero sample means could be used to extract a long run trend that might be less sensitive to the sample period employed.

The model is set up in a state-space framework, well known from time series modeling. This makes the use of the Kalman filtering technique possible, still allowing econometric estimation and prediction in a rather straightforward way, as in the Girosi and King (2005) reformulation of the Lee and Carter (1992)-model.

As an application we investigate the modeling of mortality in the Netherlands. The available data range from 1853 to 2003 and contain two peaks: one in 1918 due to the "Spanish flu" and one in 1945 due to the "Dutch Hunger Winter" at the end of the Second World War. We use the latter peak to illustrate the impact of working with a time-varying drift instead of a constant one.

The remainder of the paper is organized as follows. In Section 2, we first provide a description of the Lee–Carter model, including the reformulation by Girosi and King (2005), and discuss some of the drawbacks of this way of modeling mortality. Section 3 introduces our approach, which we illustrate in Section 4 using Dutch data on mortality. Section 5 concludes.

2. The Lee and Carter approach

In this section we first describe the Lee and Carter (1992)-approach and the way it is usually estimated. Then we present the reformulation presented in Girosi and King (2005) and we present the forecasting of future mortality based on the reformulation by Girosi and King (2005), together with some of the limitations of this way of modeling and forecasting mortality. In the next section we then introduce our alternative approach.

Let D_{xt} be the number of people with age x that died in year t, and E_{xt}, the exposure being the number of person years[1] with age x in year t, with $x \in \{1, \ldots, na\}$, and $t \in \{1, \ldots, T\}$. We consider modelling of[2]

$$m_{xt} = \ln\left(\frac{D_{xt}}{E_{xt}}\right). \tag{1}$$

Define

$$m_t = \begin{pmatrix} m_{1,t} \\ \vdots \\ m_{na,t} \end{pmatrix}, \tag{2}$$

[1] For more details on the definition and the estimation of E_{xt}, see Gerber (1997).

[2] Lee and Carter (1992) use the log of the *central death rate*. The central death rate of an individual with age x at time t is defined as a weighted average of the number of deaths in periods $t-1, t, t+1$, divided by a weighted average of the exposure for individuals with age $x-2, \ldots, x+2$. For additional details, see Benjamin and Pollard (1993).

then the model according to Lee and Carter (1992) can be formulated as

$$m_t = \alpha + \beta\gamma_t + \delta_t, \tag{3}$$

with unknown parameter vectors $\alpha = (\alpha_1, \ldots, \alpha_{na})'$ and $\beta = (\beta_1, \ldots, \beta_{na})'$, and a vector of (measurement) error terms $\delta_t = \left(\delta_{1,t}, \ldots, \delta_{na,t}\right)'$, where

$$\{\gamma_t\}_{t=1}^{T}$$

is a one-dimensional underlying latent process, assumed to be governed by

$$\gamma_t = c_0 + c_1\gamma_{t-1} + \cdots + c_k\gamma_{t-k} + \epsilon_t, \tag{4}$$

with unknown parameters $c_0, c_1, \ldots, c_k$, and error term ϵ_t satisfying

$$\epsilon_t = \omega_t + d_1\omega_{t-1} + \cdots + d_\ell\omega_{t-\ell}, \tag{5}$$

with unknown parameters $d_1, \ldots, d_\ell$, where the error term ω_t and error term vector δ_t are white noise, satisfying the distributional assumption

$$\begin{pmatrix}\delta_t\\ \omega_t\end{pmatrix}\Bigg|\,\mathcal{F}_{t-1} \sim \left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_\delta & 0\\ 0 & \sigma_\omega^2\end{pmatrix}\right),$$

with $\mathcal{F}_{t-1}$ representing the information up to time $t-1$, and with Σ_δ the unknown covariance matrix of δ_t and σ_ω^2 the unknown variance of ω_t. The error term ω_t driving the γ_t process is assumed to be uncorrelated with the vector of error terms δ_t appearing in the m_t equation.

As originally proposed by Lee and Carter (1992), the model is usually estimated in several steps. In the first step, Singular Value Decomposition (SVD) is applied to retrieve the underlying latent process, yielding $\{\widehat{\gamma}_t\}_{t=1}^{T}$. Secondly, OLS regressions are run for each age group $x = 1, \ldots, na$, to estimate the age-specific parameters, resulting in $\widehat{\alpha}$ and $\widehat{\beta}$. Thirdly, the estimated $\{\widehat{\gamma}_t\}_{t=1}^{T}$ is adjusted to ensure equality between the observed and model-implied number of deaths in a certain period, i.e., $\{\widehat{\gamma}_t\}_{t=1}^{T}$ is replaced by $\{\widetilde{\gamma}_t\}_{t=1}^{T}$ such that:

$$\sum_{x=1}^{na} D_{xt} = \sum_{x=1}^{na}\left[E_{xt}\exp(\widehat{\alpha}_x + \widehat{\beta}_x\widetilde{\gamma}_t)\right]. \tag{6}$$

Finally, the Box–Jenkins method is used to identify and estimate the dynamics of the latent factor $\widetilde{\gamma}_t$.[3]

Typically, when estimating the Lee and Carter (1992)-model, one usually infers that

$$c_0 = c, \qquad c_1 = 1, \qquad c_2 = c_3 = \cdots = 0,$$
$$d_1 = d_2 = \cdots = 0,$$

meaning that the underlying latent process is a random walk with drift. Thus, the typical version of the Lee and Carter (1992)-model, that is estimated and applied in forecasting, is given by

$$m_t = \alpha + \beta\gamma_t + \delta_t, \tag{7}$$

with

$$\gamma_t = c + \gamma_{t-1} + \epsilon_t, \tag{8}$$

where

$$\begin{pmatrix}\delta_t\\ \epsilon_t\end{pmatrix}\Bigg|\,\mathcal{F}_{t-1} \sim \left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_\delta & 0\\ 0 & \sigma_\epsilon^2\end{pmatrix}\right).$$

Following Girosi and King (2005) we can rewrite this version of the Lee and Carter (1992)-model, yielding

$$m_t = \alpha + \beta\gamma_t + \delta_t \tag{9}$$
$$= \beta c + (\alpha + \beta\gamma_{t-1} + \delta_{t-1}) + (\beta\epsilon_t + \delta_t - \delta_{t-1}) \tag{10}$$
$$= \theta + m_{t-1} + \zeta_t \tag{11}$$

with

$$\theta = \beta c, \qquad \zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}.$$

As noted by Girosi and King (2005), the typical Lee and Carter (1992)-model rewritten in this way, can easily be estimated and predicted. Indeed, with $\Delta m_t = m_t - m_{t-1}$, we can estimate θ simply by the time average of Δm_t, i.e., by

$$\widehat{\theta}_T = \frac{1}{T-1}\sum_{t=2}^{T}\Delta m_t = \frac{1}{T-1}\left(m_T - m_1\right). \tag{12}$$

This estimator has well known (T-asymptotic) characteristics. Predictions of future values of $m_{T+\tau}$, for $\tau = 1, 2, \ldots$, as well as the construction of the corresponding prediction intervals, can be based upon

$$m_{T+\tau} = m_T + \theta\tau + \sum_{t=T+1}^{T+\tau}\zeta_t. \tag{13}$$

For instance, Girosi and King (2005), ignoring the moving average character of the error terms ζ_t, construct as predictors of $m_{T+\tau}$

$$\widehat{E}\left(m_{T+\tau}|\mathcal{F}_T\right) = m_T + \widehat{\theta}_T\tau. \tag{14}$$

Thus, as prediction for a particular age (group) $x \in \{1, \ldots, na\}$, one can simply take the straight line going through the corresponding components of m_1 and m_T, extrapolated into the future.

To deal with the potential moving average character of the error term ζ_t, one could maintain the structure $\zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}$ following from Lee and Carter (1992), or, alternatively, one could postulate that ζ_t follows an MA(1) structure given by

$$\zeta_t = \xi_t + \Theta\xi_{t-1}, \tag{15}$$

with Θ an $(na \times na)$ matrix of unknown parameters, and where ξ_t is an na-dimensional vector of white noise, satisfying the

[3] The readjustment of the latent process in the third step is done in order to avoid sizeable differences between the observed and the model-implied number of deaths. Other advantages of the readjustment have been mentioned in Lee (2000). However, the fact that the readjustment is done without re-estimating the age-specific sensitivity parameters $\widehat{\beta}_x$ also has several drawbacks. First, since the estimated variables $\widehat{\gamma}_t$, obtained in the first step, are adjusted after the age-specific coefficients in the OLS regressions are estimated, the resulting term $\widehat{\beta}_x\widetilde{\gamma}_t$ might not accurately describe the movements in the log death rates $m_{x,t}$ anymore. Second, the standard error estimated for $\widehat{\beta}_x$ in the age-specific regressions does not necessarily describe the correct size of the uncertainty for the parameter estimates.

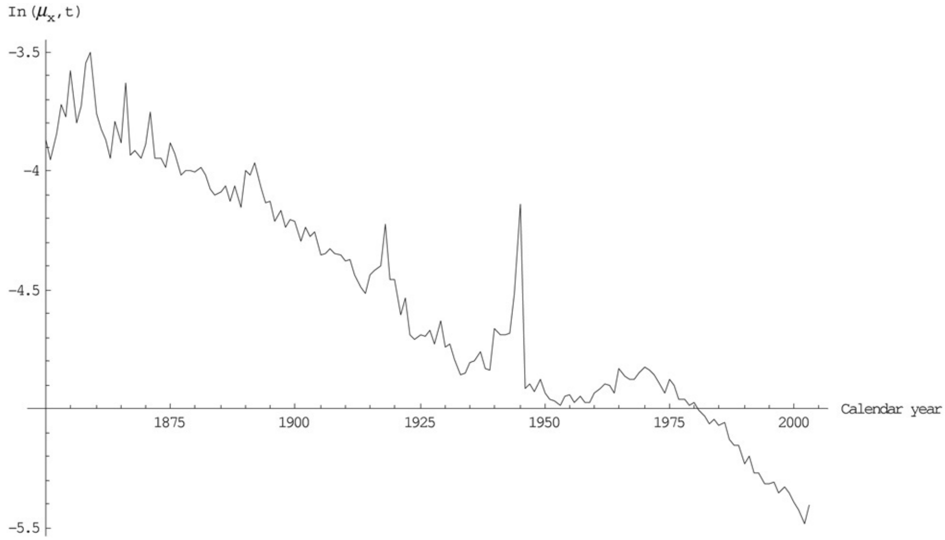

Fig. 1. Log mortality for the age group of 50–54, men. The figure shows log mortality data of Dutch men for the age group of 50–54 years during the sample period 1850–2003. Data source: Human Mortality Database.

distributional assumptions

$$\xi_t|\mathcal{F}_{t-1} \sim \left(0, \Sigma_\xi\right).$$

With these modifications, the Lee and Carter (1992)-model becomes

$$m_t = \theta + m_{t-1} + \zeta_t, \tag{16}$$

$$\zeta_t = \xi_t + \Theta\xi_{t-1}, \tag{17}$$

$$\xi_t|\mathcal{F}_{t-1} \sim \left(0, \Sigma_\xi\right).$$

A main drawback of the Lee and Carter (1992)-model follows from the Girosi and King (2005) specification. Ignoring for simplicity the possible forecast correction due to an MA error term (which only affects the level but not the slope), the forecast of age (group) $x \in \{1, \ldots, na\}$ is essentially the straight line through $m_{x,1}$ and $m_{x,T}$, extrapolated into the future. Fig. 1 shows Dutch mortality data of the age group 50–54 years during the sample period 1850–2003. As this figure illustrates, the mortality data is rather volatile, particularly at the beginning of the sample period, but also around the first and second world wars. This means that the estimates, and, thus, the mortality forecasts, might be rather sensitive to the exact sample period used in estimation: The straight lines through $m_{x,t}$ and $m_{x,\tau}$ may be different for different values of t or τ, resulting in quite different long run forecasts.

In the next section we present an extension of the Lee and Carter (1992)-model, starting from the Girosi and King (2005)-reformulation, that is aimed to result in estimates of the long run trend that might be less sensitive to the particular sample period employed.

3. Lee–Carter with time-varying drift

In this section we present our generalization of the Lee and Carter (1992)-approach, taking as starting point the version typically found in empirical studies, as described in the previous section. We first describe the model and its motivation, then its estimation and its use in forecasting.

As generalization of the Lee and Carter (1992)-approach, we propose the following model for m_t:

$$m_t = \theta_t + m_{t-1} + \zeta_t \tag{18}$$

with

$$\theta_t = a + Bu_t \tag{19}$$

$$u_t - \mu_u = \Gamma\left(u_{t-1} - \mu_u\right) + \eta_t \tag{20}$$

with u_t an nf-dimensional vector of underlying latent factors, driving the "constant" θ_t, where $a \in \mathbb{R}^{na}$, $B \in \mathbb{R}^{na\times nf}$, $\mu_u = E\left(u_t\right) \in \mathbb{R}^{nf}$, and $\Gamma \in \mathbb{R}^{nf\times nf}$ are unknown parameter vectors and matrices, where the na-dimensional vector of (measurement) errors ζ_t satisfies

$$\zeta_t = \xi_t + \Theta\xi_{t-1} \tag{21}$$

with $\Theta \in \mathbb{R}^{na\times na}$ a matrix with unknown parameters capturing the MA effects, where the nf-dimensional vector of (measurement) errors η_t satisfies

$$\eta_t = \psi_t + \Xi\psi_{t-1} \tag{22}$$

with $\Xi \in \mathbb{R}^{nf\times nf}$ a matrix with unknown parameters capturing the MA effects, and where the vectors of error terms ψ_t and ξ_t are white noise, satisfying the distributional assumption

$$\left.\begin{pmatrix}\psi_t\\ \xi_t\end{pmatrix}\right| \mathcal{F}_{t-1} \sim \left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_\psi & 0\\ 0 & \Sigma_\xi\end{pmatrix}\right)$$

with $\Sigma_\psi \in \mathbb{R}^{nf\times nf}$ and $\Sigma_\xi \in \mathbb{R}^{na\times na}$ the unknown covariance matrices of ψ_t and ξ_t, respectively.[4] The vectors of error terms ψ_t and ζ_t are assumed to be uncorrelated. The Lee and Carter (1992)-model is obtained as a special case by imposing $B = 0$.

[4] This model can easily be generalized to allow for a higher order moving average in the equation for m_t or a higher order autoregression and higher order moving average in the equation for u_t, but for the sake of simplicity (and also based on empirical outcomes), we shall focus in the sequel on the current special version of the model. However, generalizing the subsequent analysis to the case with (higher order) moving average terms is straightforward.

The long run mean of the u_t process is postulated to be equal to zero, i.e., $\mu_u = 0$, so that u_t corresponds to a process that fluctuates around zero. The changes in the drift term θ_t is postulated to be picked up by these changes in u_t. To capture comovements between different age groups, the time-varying short run changes are modeled as Bu_t, with u_t low dimensional and with B the age- (group-)specific sensitivities to the underlying time-varying latent process.

To estimate the model, we apply the state-space method combined with the Kalman filtering technique (see, for instance, Durbin and Koopman (2001), Hamilton (1994)). The model can straightforwardly be put in a state-space form, with as 'observation equation'

$$\begin{aligned}\Delta m_t &\equiv m_t - m_{t-1} \\ &= (a + B\mu_u) + \left[B \ \vdots \ 0 \ \vdots \ 0 \ \vdots \ I \ \vdots \ \Xi\right] \\ &\times \begin{pmatrix} u_t - \mu_u \\ \psi_t \\ \psi_{t-1} \\ \xi_t \\ \xi_{t-1} \end{pmatrix} \equiv A + H' z_t,\end{aligned} \tag{23}$$

where we define

$$A \equiv a + B\mu_u, \quad \text{and} \quad H \equiv \left[B \ \vdots \ 0 \ \vdots \ 0 \ \vdots \ I \ \vdots \ \Xi\right],$$

and with 'state equation'

$$\begin{aligned} z_t &\equiv \begin{pmatrix} u_t - \mu_u \\ \psi_t \\ \psi_{t-1} \\ \xi_t \\ \xi_{t-1} \end{pmatrix} \\ &= \begin{pmatrix} \Gamma & \Xi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{pmatrix} \begin{pmatrix} u_{t-1} - \mu_u \\ \psi_{t-1} \\ \psi_{t-2} \\ \xi_{t-1} \\ \xi_{t-2} \end{pmatrix} + \begin{pmatrix} \psi_t \\ \psi_t \\ 0 \\ \xi_t \\ 0 \end{pmatrix} \\ &\equiv F z_{t-1} + v_t, \end{aligned} \tag{24}$$

where

$$F \equiv \begin{pmatrix} \Gamma & \Xi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{pmatrix} \quad \text{and} \quad v_t \equiv \begin{pmatrix} \psi_t \\ \psi_t \\ 0 \\ \xi_t \\ 0 \end{pmatrix}.$$

Let Σ_v denote the covariance matrix of v_t. Following Hamilton (1994), we can easily derive the likelihood function corresponding to our state-space model specification, under the additional assumption that the distribution of

$$\left.\begin{pmatrix} \psi_t \\ \xi_t \end{pmatrix}\right| \mathcal{F}_{t-1}$$

is normal. Let $\widehat{z}_{t|t-1}$ denote the best linear estimate of z_t, given information available up to time $t-1$, and let $P_{t|t-1}$ denote the forecast error covariance matrix, i.e.,

$$\widehat{z}_{t|t-1} = E\left(z_t \mid \mathcal{F}_{t-1}\right) = F z_{t-1} \tag{25}$$

$$P_{t|t-1} = E\left[\left(z_t - \widehat{z}_{t|t-1}\right)\left(z_t - \widehat{z}_{t|t-1}\right)' \mid \mathcal{F}_{t-1}\right] = \Sigma_v, \tag{26}$$

where

$$\widehat{z}_{1|0} = E\left(z_1\right) = 0, \qquad P_{1|0} = E\left(z_1 z_1'\right) = \Sigma_v.$$

Then, if the initial state and the innovations are multivariate Gaussian, the distribution of Δm_t, conditional on the information $\mathcal{F}_{t-1}$ up to time $t-1$ is normal:

$$\Delta m_t \mid \mathcal{F}_{t-1} \sim N\left(A + H'\widehat{z}_{t|t-1}, H' P_{t|t-1} H\right), \tag{27}$$

and the likelihood contribution for observation t is given by:

$$\begin{aligned} L_t &= (2\pi)^{-\frac{m}{2}} * \det\left(H' P_{t|t-1} H\right)^{-\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2}\left(\Delta m_t - \left(A + H'\widehat{z}_{t|t-1}\right)\right)' \left(H' P_{t|t-1} H\right)^{-1} \right. \\ &\times \left. \left(\Delta m_t - \left(A + H'\widehat{z}_{t|t-1}\right)\right)\right). \end{aligned} \tag{28}$$

In the sequel we shall consider the likelihood as a quasi-likelihood, and calculate the asymptotic covariance matrix of the quasi-maximum likelihood accordingly, to allow for the possibility of non-normally distributed error terms, thus assuming only that the first moments are correctly specified.

The construction of predictions as well as the corresponding prediction intervals can be based upon

$$m_{T+\tau} = m_T + \sum_{t=T+1}^{T+\tau} \left(\theta_t + \zeta_t\right) \tag{29}$$

$$= m_T + \sum_{t=T+1}^{T+\tau} \left(a + Bu_t + \xi_t + \Theta\xi_{t-1}\right) \tag{30}$$

$$= m_T + a\tau + B\sum_{t=T+1}^{T+\tau} u_t + \sum_{t=T+1}^{T+\tau} \left(\xi_t + \Theta\xi_{t-1}\right). \tag{31}$$

As prediction of future values of $m_{t+\tau}$ we shall use

$$\widehat{E}\left(m_{T+\tau}|\mathcal{F}_T\right) = m_T + \widehat{a}\tau + \widehat{B}\sum_{t=T+1}^{T+\tau} \widehat{E}\left(u_t|\mathcal{F}_T\right) + \widehat{\Theta}\widehat{\xi}_T, \tag{32}$$

with

$$\begin{aligned}\widehat{E}\left(u_{T+t}|\mathcal{F}_T\right) &= \left(\widehat{\mu}_u + \widehat{\Gamma}\widehat{\mu}_u + \cdots + \widehat{\Gamma}^{t-1}\widehat{\mu}_u\right) \\ &\quad + \widehat{\Gamma}^t\left(\widehat{u}_T - \widehat{\mu}_u\right) + \widehat{\Gamma}^{t-1}\widehat{\Xi}\widehat{\psi}_T,\end{aligned} \tag{33}$$

where the hats ($\widehat{\ }$) indicate estimated values of the corresponding parameters. When calculating the prediction intervals we make the assumption that the white noise processes follow a normal distribution. These prediction intervals are constructed via simulation.[5] Most estimators follow straightforwardly from

[5] Constructing the prediction intervals via simulation can be done easily. Given the asymptotic distribution of the quasi-maximum likelihood estimator,

maximizing the quasi-log-likelihood. Notice that under the identifying assumption $\mu_u = 0$, it makes sense to take $\widehat{\mu}_u = 0$ and $\widehat{a} = \widehat{A}$ (using $A = a + B\mu_u$). Thus, imposing $\mu_u = 0$, we can estimate the long run trend (a) by the estimated sample trend ($\widehat{A}$). However, we can also obtain as alternative estimator for μ_u

$$\widehat{\mu}_u = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t = \frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t. \tag{34}$$

In finite samples this alternative estimator's deviation from zero might reflect the model's difficulty in estimating the long run trend a. Instead of estimating the long run trend by the sample trend $\widehat{a} = \widehat{A}$ (when $\mu_u = 0$), we might then alternatively estimate the long run trend by

$$\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u, \tag{35}$$

using the alternative estimator for μ_u. In this way we might obtain an estimate of the long run trend that is somewhat less sensitive to volatility in the mortality data than the estimator $\widehat{a} = \widehat{A}$. In the empirical analysis of the next section we shall investigate both ways of estimating the long run trend.

4. Empirical analysis

4.1. Data

We use 154 yearly observations of age-specific death numbers (D_{xt}) and exposures (E_{xt}) for men in the Netherlands, from 1850 till 2003, provided by The Human Mortality Database.[6] As in Lee and Carter (1992), we create the following age groups: 1–4, 5–9, 10–14, ..., 80–84, and 85+. Since the database provides data starting at the middle of the 19th century, and the number of people in age groups above 85 (for example, 85–89, or 90–94, etc.) is relatively low in that period, we merge all the age groups above 85, resulting in the 85+ category.[7] This results in a total of 18 age groups, so that in our empirical application we have $na = 18$.

4.2. Further specifications

In the empirical application, we impose some additional structure. For identification purposes we impose that the covariance matrix of the vector of error terms ψ_t is the identity matrix of dimension $nf \times nf$, and that the matrices Γ and Ξ appearing in the equation for the underlying latent vector u_t are restricted to be lower triangular matrices. In addition, we impose extra structure for the covariance matrix of the vector of error terms ξ_t, and for the moving average matrix Θ[8]:

$$\Sigma_\xi = \mathrm{diag}\left(\sigma^2_{\xi,1}, \ldots, \sigma^2_{\xi,na}\right), \tag{36}$$

$$\Theta = \mathrm{diag}\left(\Theta_1, \ldots, \Theta_{na}\right). \tag{37}$$

Moreover, with na age groups and nf latent factors, the number of parameters in a and B to be estimated equals $na \times (nf + 1)$. In order to reduce this number of parameters, and to avoid localized age-induced anomalies, we use spline interpolation (see also Renshaw and Haberman (2003a)), described in Appendix.

4.3. Sample period sensitivity

Before presenting the estimation results when using the whole sample, we first illustrate the effect of changing the sample period. In Table 1 we present for some age groups in the range 40–69 years the estimation results for two versions of the model, namely, one with a single latent factor following an AR(1)-process (1F AR) and one with a single factor following an MA(1)-process (1F MA), where the sample is either 1850–1945 or 1850–1946, where the two end years have substantially different mortality data, due to the peak in the registered number of deaths in the year 1945. We take these two subsamples for illustrative purposes only. Comparing other selected subsamples would result in similar, but less pronounced findings. We report the estimation results for $\widehat{a} = \widehat{A}$ (under the heading A) and for $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u (under the heading a). We also present the estimates of θ according to the Lee and Carter (1992) specification following Girosi and King (2005) (reported under the heading A).

Changing the sample period from 1850–1945 to 1850–1946 has a dramatic impact on the estimates of A, particularly in case of the Lee and Carter (1992) specification. However, the corresponding estimates in terms of $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ are much more stable, indicating that this way of estimating the long run trend indeed does seem to do its job: making the estimated long run trend less sensitive to shocks in the data.

Since the estimates of a depend on the sample estimates of μ_u, in Fig. 2 we further illustrate the sensitivity of $\widehat{\mu}_u$ to the sample period employed. In the upper panels of Fig. 2 we consider the 1F AR and the 1F MA models for the sample between 1850 and 1945 and report how the sample estimates of μ_u

$$\widehat{\mu}_u = \frac{1}{T - t + 1}\sum_{t=1850}^{T}\widehat{u}_t \tag{38}$$

evolve if T runs from 1875 till 1945, based on the estimated $\widehat{u}_t$-s. This shows the impact of the positive shock in the year 1945: $\widehat{\mu}_u$ estimated at the end of this sample period (denoted

we can simulate parameter values, and given these simulated parameter values, we can simulate the process for $m_{T+\tau}$ making use of the assumption that the white noise processes follow a normal distribution. In this way we can generate both prediction intervals, conditional upon given parameter estimates, and prediction intervals also capturing estimation inaccuracy.

[6] Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at http://www.mortality.org or http://www.humanmortality.de (data downloaded on 01.12.2004).

[7] Alternatively, assumptions on old age mortality could be imposed (see for example, Coale and Guo (1989)).

[8] This additional structure imposed on Σ_ξ and Θ is also intended to ease the estimation, since this structure actually implies some overidentification constraints.

Table 1
Estimation results of A-s and a-s for various models and age groups, using as subsamples 1850–1945 and 1850–1946

Age group (x)	1850–1945			1850–1946		
	LC	1F AR	1F MA	LC	1F AR	1F MA
	A_x			A_x		
45–49	−0.0019	−0.0027	−0.0032	−0.0121	−0.0082	−0.0089
50–54	−0.0028	−0.0033	−0.0037	−0.0109	−0.0078	−0.0084
55–59	−0.0016	−0.0024	−0.0027	−0.0091	−0.0063	−0.0069
60–64	−0.0011	−0.0017	−0.0021	−0.0075	−0.0052	−0.0057
65–69	−0.0003	−0.0010	−0.0013	−0.0066	−0.0043	−0.0048
		a_x			a_x	
45–49		−0.0034	−0.0044		−0.0046	−0.0061
50–54		−0.0039	−0.0049		−0.0047	−0.0060
55–59		−0.0029	−0.0037		−0.0037	−0.0048
60–64		−0.0022	−0.0030		−0.0028	−0.0038
65–69		−0.0014	−0.0021		−0.0021	−0.0030

The table shows the impact of the sample period (1850–1945, or alternatively 1850–1946) on the differences between the long run trend ($\widehat{a}$) and sample trend ($\widehat{A}$) estimated by various models: Lee–Carter (LC), single-factor first-order autoregressive (1F AR) and single-factor first-order moving average (1F MA), for various age groups: 45–49, 50–54, etc.

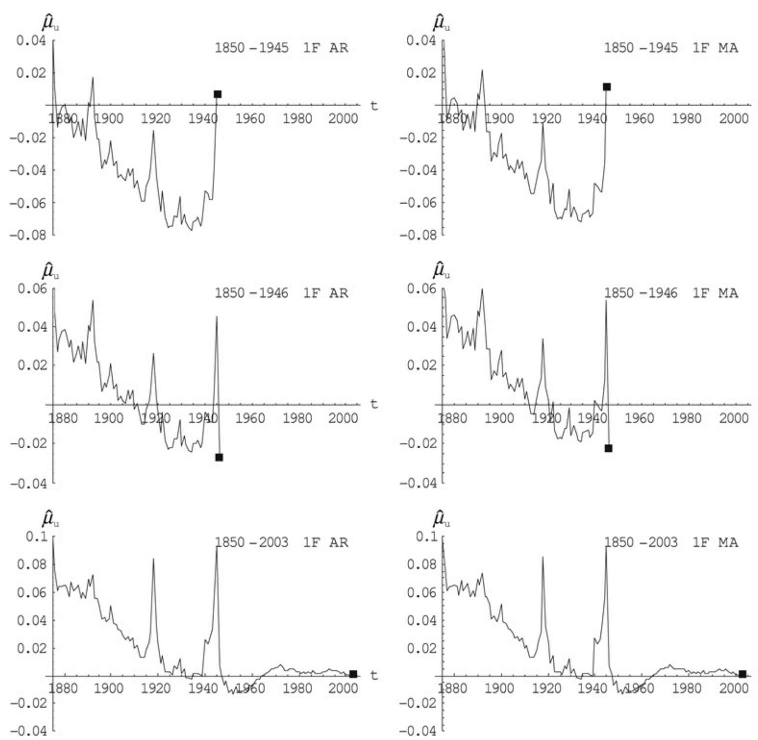

Fig. 2. Sensitivity of $\widehat{\mu}_u$ to various sample periods. The figure shows the sensitivity of the estimated mean of the latent process ($\widehat{\mu}_u$, represented by a square at the end of the employed sample period on the figure) to different sample periods: 1850–1945, 1850–1946, and 1850–2003, with two model specifications: single-factor first-order autoregressive (1F AR) and single-factor first-order moving average (1F MA).

by the square on the figure) deviates positively from zero for both model specifications.

Similarly, the middle panels show the outcome when the 1F AR and the 1F MA specifications are estimated for the sample period 1850–1946, with a large negative shock in the registered number of deaths in year 1946. The end of sample estimates for μ_u for both model specifications are now negative.

Finally, we consider the full sample between 1850 and 2003 in the lower panels of Fig. 2 without any (visible) shock in year 2003. $\widehat{\mu}_u$ is close to zero at the end of the sample period, both for 1F AR and 1F MA, which indicates that the estimated sample trend ($\widehat{A}$) seems to be sufficiently close to the long run trend estimate ($\widehat{a}$).

4.4. Estimation results

In this subsection we present the estimation results of the various model specifications using the whole available sample period.[9] We start with the Lee and Carter (1992)-benchmark, following the specification of Girosi and King

[9] In Fig. 1 describing the log mortality of men for the age group of 50–54 over time, two events that resulted in an increase of the registered number of

Table 2
One-factor LC model with splines

Age group (x)	Coefficients		
	A_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1–4	−0.032 (0.015)	−0.007 (0.524)	0.152 (0.014)
5–9	−0.024 (0.011)	−0.030 (0.156)	0.161 (0.014)
10–14	−0.021 (0.012)	−0.162 (0.190)	0.167 (0.019)
15–19	−0.020 (0.012)	−0.084 (0.201)	0.172 (0.032)
20–24	−0.019 (0.012)	−0.335 (0.229)	0.219 (0.038)
25–29	−0.018 (0.011)	−0.353 (0.307)	0.211 (0.036)
30–34	−0.017 (0.010)	−0.387 (0.304)	0.196 (0.033)
35–39	−0.015 (0.008)	−0.360 (0.330)	0.167 (0.029)
40–44	−0.013 (0.007)	−0.400 (0.336)	0.145 (0.022)
45–49	−0.012 (0.005)	−0.396 (0.374)	0.117 (0.019)
50–54	−0.010 (0.004)	−0.490 (0.270)	0.096 (0.013)
55–59	−0.009 (0.004)	−0.501 (0.287)	0.084 (0.012)
60–64	−0.007 (0.003)	−0.452 (0.229)	0.078 (0.011)
65–69	−0.006 (0.003)	−0.506 (0.155)	0.073 (0.010)
70–74	−0.005 (0.003)	−0.511 (0.156)	0.069 (0.008)
75–79	−0.004 (0.002)	−0.598 (0.152)	0.073 (0.008)
80–84	−0.003 (0.002)	−0.645 (0.142)	0.073 (0.006)
85+	−0.002 (0.002)	−0.684 (0.176)	0.080 (0.006)
Log-likelihood	1944.46		

The table shows the estimated parameters of the Lee–Carter model.
This table reports QML estimates and standard errors of the one-factor LC mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

(2005). Furthermore, we estimate the model described in Section 3 for different specifications of the latent factor. We consider one- and two-factor versions, following first-order autoregressive (AR) or moving average (MA) processes, and we estimate the long run trend both by means of $\widehat{a} = \widehat{A}$ and by means of $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u.

Table 2 contains the Lee and Carter (1992)-benchmark.[10] To ease a comparison with the other models we use the heading A to refer to the parameter vector θ. The results indicate that the decreasing trend in mortality is steepest for the youngest age group, increasing to a value close to zero (and statistically insignificant) for the oldest age group. The moving average terms are insignificant for the lower age groups, but become significantly negative for the older age groups. The standard deviations of the white noise error terms ξ_t are always substantial and estimated quite accurately, with the higher age groups having smaller standard deviations. These standard deviations are larger for the younger age groups, particularly, for the age groups 20–29 year.

deaths can be identified: (1) the "Spanish flu" epidemic around the year 1918, and, (2) the so-called "Dutch Hunger Winter" at the end of the Second World War. These two events affected all age groups, some of them to a larger, some of them to a smaller extent. The main reason why we did not include dummies into the time series of the log mortality rates to filter out these events as happens sometimes in other studies is as follows. The two events increased the number of deaths, particularly among the more vulnerable, with as consequence that in the subsequent periods the mortality experience of a potentially stronger population is observed with better survival characteristics. If we had filtered out the effects of the two events, the properties of the time series process would be potentially affected, since it reflected the mortality experience as if the population consisted of the stronger members only.

[10] There is hardly any difference between the original Lee and Carter (1992)-results without using splines and the results with splines that we employ. The restrictions imposed by the six parameter spline only reduces the log-likelihood value very marginally while the age profile is slightly smoothed.

In Tables 3 and 4 we present the estimation results for the one-factor models, with Table 3 containing the estimation results in case the latent factor follows an AR(1) process, while in Table 4 the latent factor follows an MA(1) process. The heading A refers to the estimate $\widehat{a} = \widehat{A}$, while the heading a refers to the estimate $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u. Compared to the Lee–Carter estimation results, we first notice a dramatic increase in the log-likelihood, suggesting a significant statistical improvement. The improved fit of these one-factor models is reflected by sometimes substantially smaller estimates of the standard deviations of the error terms ξ_t compared to the Lee–Carter specifications. In addition, the structure of the error terms ξ_t also changes. Particularly, the moving average coefficients (Θ) of the lower age groups get more negative values and become significantly different from zero.

The estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ are quite comparable to the ones according to $\widehat{A}$. So, the model seems to be able to fit the long run trend reasonably well. This is also reflected in the lower panels of Fig. 2. However, the long run trend, as estimated by $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, is estimated less accurately. The factor loadings (B) turn out to be very significant, revealing a hump shape, with a peak at the age groups 20–29 years. In the AR version, the autoregression coefficient in the underlying latent process turns out to be insignificant. The same applies to the moving average term in the MA specification.

Based on the Ljung–Box test the residuals of both the AR and the MA specifications seem to have the characteristics of a white noise. So, from this perspective there seems to be no need to include an ARMA specification.

In Tables 5 and 6 we present the estimation results for the two-factor models, with Table 5 containing the AR(1) estimation results, and Table 6 the MA(1) results. In both cases the log-likelihood increases substantially compared to the corresponding one-factor cases, but for the MA case slightly more than for the AR case. A likelihood ratio test reveals that the one-factor versions are rejected against the two-factor variants. However, the long run trend estimates remain more or less the same as in case of the one-factor models, also in terms of their estimation accuracy, and the same applies to the moving average terms Θ. The standard deviations of the white noise error terms ξ_t decrease slightly, reflecting the better fit of the two-factor variant. In the AR version, the autoregression coefficient of the first latent factor turns out to be significant. The same applies to the moving average term of the first latent factor in the MA specification. Similarly to the one-factor case, the residuals of the two-factor AR and MA latent processes seem to be white noise, so we did not include the ARMA specification.

The two factors can already capture most of the common properties (correlation) among separate age groups in the Netherlands. Both for the AR and the MA specifications, the first factor seems to be responsible for driving the old age mortality, taking into account significant estimates of the age groups above the age of 50. We could call this factor the *old age* factor. The second factor seems to drive the young and middle age mortality, since it affects mostly the middle-aged groups,

Table 3
One-factor model, AR latent factor

	Coefficients				
Γ	−0.244				
	(0.240)				
Σ_ψ	*1*				
Age group (x)	A_x	a_x	B_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1–4	−0.032 (0.009)	−0.032 (0.016)	0.114 (0.030)	−0.500 (0.231)	0.120 (0.009)
5–9	−0.024 (0.008)	−0.024 (0.016)	0.115 (0.020)	−0.358 (0.185)	0.116 (0.008)
10–14	−0.021 (0.010)	−0.021 (0.019)	0.143 (0.024)	−0.608 (0.121)	0.097 (0.008)
15–19	−0.020 (0.012)	−0.020 (0.023)	0.178 (0.033)	−0.467 (0.100)	0.086 (0.006)
20–24	−0.020 (0.013)	−0.019 (0.026)	0.198 (0.040)	−0.454 (0.159)	0.108 (0.017)
25–29	−0.019 (0.013)	−0.019 (0.026)	0.199 (0.042)	−0.590 (0.075)	0.077 (0.008)
30–34	−0.017 (0.012)	−0.017 (0.024)	0.185 (0.039)	−0.659 (0.056)	0.066 (0.007)
35–39	−0.016 (0.010)	−0.016 (0.021)	0.162 (0.033)	−0.723 (0.075)	0.044 (0.004)
40–44	−0.014 (0.009)	−0.014 (0.018)	0.136 (0.027)	−0.767 (0.051)	0.036 (0.002)
45–49	−0.012 (0.007)	−0.012 (0.015)	0.112 (0.022)	−0.492 (0.087)	0.035 (0.003)
50–54	−0.011 (0.006)	−0.011 (0.012)	0.093 (0.019)	−0.390 (0.097)	0.041 (0.004)
55–59	−0.009 (0.005)	−0.009 (0.010)	0.078 (0.017)	−0.362 (0.092)	0.039 (0.004)
60–64	−0.008 (0.005)	−0.008 (0.009)	0.067 (0.016)	−0.307 (0.102)	0.042 (0.004)
65–69	−0.006 (0.004)	−0.006 (0.008)	0.058 (0.015)	−0.375 (0.116)	0.044 (0.004)
70–74	−0.005 (0.004)	−0.005 (0.007)	0.052 (0.014)	−0.478 (0.112)	0.047 (0.004)
75–79	−0.004 (0.003)	−0.004 (0.006)	0.048 (0.013)	−0.700 (0.110)	0.051 (0.005)
80–84	−0.003 (0.003)	−0.003 (0.006)	0.044 (0.012)	−0.813 (0.077)	0.057 (0.005)
85+	−0.002 (0.003)	−0.002 (0.006)	0.041 (0.012)	−0.792 (0.100)	0.067 (0.005)
Log-likelihood	3515.12				

The table shows the estimated parameters of the model driven by a single-factor autoregressive latent process.
This table reports QML estimates and standard errors of the one-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 4
One-factor model, MA latent factor

	Coefficients				
Ξ	−0.326				
	(0.385)				
Σ_ψ	*1*				
Age group (x)	A_x	a_x	B_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1–4	−0.032 (0.008)	−0.032 (0.014)	0.115 (0.031)	−0.499 (0.235)	0.121 (0.010)
5–9	−0.024 (0.007)	−0.024 (0.014)	0.114 (0.020)	−0.357 (0.187)	0.116 (0.008)
10–14	−0.020 (0.008)	−0.021 (0.017)	0.142 (0.022)	−0.607 (0.123)	0.097 (0.008)
15–19	−0.020 (0.010)	−0.020 (0.020)	0.175 (0.031)	−0.466 (0.101)	0.086 (0.006)
20–24	−0.019 (0.011)	−0.020 (0.022)	0.195 (0.037)	−0.454 (0.158)	0.108 (0.017)
25–29	−0.018 (0.011)	−0.019 (0.022)	0.196 (0.038)	−0.589 (0.075)	0.077 (0.008)
30–34	−0.017 (0.010)	−0.017 (0.021)	0.182 (0.036)	−0.658 (0.056)	0.066 (0.007)
35–39	−0.015 (0.009)	−0.016 (0.018)	0.160 (0.030)	−0.722 (0.075)	0.044 (0.004)
40–44	−0.014 (0.007)	−0.014 (0.015)	0.134 (0.025)	−0.768 (0.052)	0.036 (0.002)
45–49	−0.012 (0.006)	−0.012 (0.013)	0.111 (0.020)	−0.493 (0.087)	0.035 (0.003)
50–54	−0.010 (0.005)	−0.011 (0.010)	0.092 (0.017)	−0.391 (0.097)	0.041 (0.004)
55–59	−0.009 (0.004)	−0.009 (0.009)	0.077 (0.016)	−0.362 (0.092)	0.039 (0.004)
60–64	−0.007 (0.004)	−0.008 (0.008)	0.066 (0.015)	−0.307 (0.102)	0.042 (0.004)
65–69	−0.006 (0.004)	−0.006 (0.007)	0.058 (0.014)	−0.375 (0.116)	0.044 (0.004)
70–74	−0.005 (0.003)	−0.005 (0.006)	0.051 (0.013)	−0.478 (0.112)	0.047 (0.004)
75–79	−0.004 (0.003)	−0.004 (0.005)	0.047 (0.013)	−0.700 (0.110)	0.051 (0.005)
80–84	−0.003 (0.003)	−0.003 (0.005)	0.043 (0.012)	−0.812 (0.076)	0.057 (0.005)
85+	−0.002 (0.003)	−0.002 (0.005)	0.040 (0.011)	−0.792 (0.100)	0.067 (0.005)
Log-likelihood	3516.43				

The table shows the estimated parameters of the model driven by a single-factor moving average latent process.
This table reports QML estimates and standard errors of the one-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

and slightly less the younger generation; however, it does not significantly influence the mortality rates of the old age groups. So, it is a *young and middle age* factor.

Table 7 compares the in-sample fit of the different models based on the cumulative sum of squared deviations of one period ahead in-sample forecasts. In case of the time-varying

Table 5
Two-factor model, AR latent factor

	Coefficients					
Γ	−0.471	*0*				
	(0.108)					
	−0.021	−0.189				
	(0.135)	(0.218)				
Σ_ψ	*1*	*0*				
	0	*1*				

Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1–4	−0.031 (0.009)	−0.032 (0.018)	0.015 (0.031)	−0.124 (0.022)	−0.576 (0.117)	0.114 (0.009)
5–9	−0.024 (0.009)	−0.024 (0.017)	0.018 (0.034)	−0.120 (0.018)	−0.403 (0.136)	0.106 (0.008)
10–14	−0.020 (0.010)	−0.021 (0.021)	0.026 (0.043)	−0.146 (0.022)	−0.629 (0.067)	0.083 (0.006)
15–19	−0.020 (0.013)	−0.020 (0.026)	0.036 (0.052)	−0.178 (0.030)	−0.508 (0.070)	0.076 (0.005)
20–24	−0.019 (0.014)	−0.019 (0.028)	0.044 (0.057)	−0.196 (0.036)	−0.555 (0.252)	0.107 (0.017)
25–29	−0.018 (0.014)	−0.018 (0.028)	0.049 (0.056)	−0.193 (0.037)	−0.688 (0.078)	0.060 (0.006)
30–34	−0.017 (0.013)	−0.017 (0.026)	0.053 (0.051)	−0.176 (0.034)	−0.673 (0.061)	0.053 (0.004)
35–39	−0.015 (0.011)	−0.015 (0.022)	0.055 (0.043)	−0.150 (0.029)	−0.734 (0.071)	0.038 (0.003)
40–44	−0.014 (0.009)	−0.014 (0.018)	0.056 (0.035)	−0.120 (0.024)	−0.738 (0.045)	0.037 (0.002)
45–49	−0.012 (0.007)	−0.012 (0.015)	0.057 (0.028)	−0.094 (0.021)	−0.539 (0.074)	0.035 (0.002)
50–54	−0.010 (0.006)	−0.010 (0.012)	0.057 (0.022)	−0.073 (0.019)	−0.478 (0.094)	0.033 (0.002)
55–59	−0.009 (0.005)	−0.009 (0.011)	0.058 (0.017)	−0.057 (0.018)	−0.466 (0.073)	0.031 (0.002)
60–64	−0.007 (0.005)	−0.007 (0.009)	0.059 (0.014)	−0.045 (0.017)	−0.400 (0.088)	0.025 (0.002)
65–69	−0.006 (0.004)	−0.006 (0.008)	0.060 (0.012)	−0.036 (0.017)	−0.584 (0.102)	0.024 (0.002)
70–74	−0.005 (0.004)	−0.005 (0.008)	0.061 (0.010)	−0.030 (0.017)	−0.556 (0.126)	0.021 (0.002)
75–79	−0.004 (0.004)	−0.004 (0.008)	0.063 (0.009)	−0.024 (0.017)	−0.518 (0.074)	0.025 (0.003)
80–84	−0.003 (0.004)	−0.003 (0.008)	0.065 (0.009)	−0.020 (0.017)	−0.466 (0.060)	0.035 (0.003)
85+	−0.002 (0.004)	−0.002 (0.008)	0.066 (0.010)	−0.015 (0.017)	−0.366 (0.121)	0.048 (0.004)
Log-likelihood	4019.59					

The table shows the estimated parameters of the model driven by a two-factor autoregressive latent process.
This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

drifts, we only report the results using the estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$. Among the models with the time-varying drift specification, the two-factor MA has the best performance, except for the youngest age groups, which is expected, taking into account the corresponding log-likelihood values. In the short run the LC model with a constant drift performs as well as the two-factor MA model in case of the older age groups, better at the young age groups, but worse in the middle age groups.

We also estimated the three-factor versions of the AR and MA models. Even though the likelihood ratio test statistic indicates that the third factor explains a significant part of the variation, the process of this third latent factor seems to follow a random walk both for the AR and the MA versions, thus, capturing essentially irregular behavior, and the factor loadings belonging to the third factor have large parameter uncertainty. This is the reason we do not report these estimates here.

4.5. Prediction

In Figs. 3–7 we plot 95%-prediction intervals for D_{xt}/E_{xt}, the ratio of age-specific death numbers and exposures, for the various models for the age group 65–69 year, taking as future the period 2003–2050. We make a distinction between a 95%-prediction interval, given the Quasi-Maximum Likelihood (QML) estimates, and a 95% prediction interval also including the estimation inaccuracy of the QML estimates. In case of the time-varying drifts, we only report for illustrative purposes the results using the estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$.

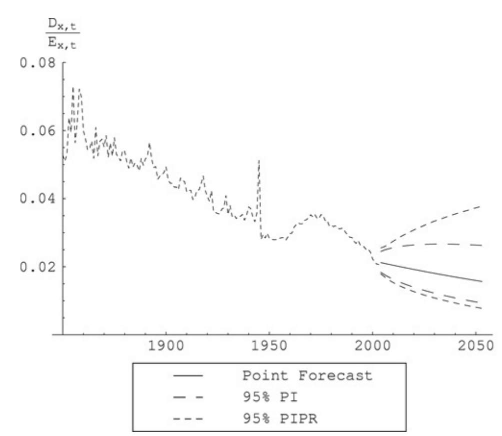

Fig. 3. Prediction age group 65–69 year; LC spline. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the LC model for the age group of 65–69 years, both without (PI) and with (PIPR) taking into account parameter estimation risk.

The prediction intervals of the various models, given the QML estimates, are quite comparable. However, as soon as we also include estimation inaccuracy in the confidence intervals, the confidence intervals become substantially wider. In case of the Lee and Carter (1992)-model this increase in prediction interval is not as large as in the other models, reflecting that in the Lee and Carter (1992) specification we use the more accurately estimated $\widehat{A}$, while in the other specifications we

Table 6
Two-factor model, MA latent factor

	Coefficients					
Ξ	−0.602	*0*				
	(0.140)					
	−0.127	−0.261				
	(0.118)	(0.358)				
Σ_ψ	*1*	*0*				
	0	*1*				
Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1–4	−0.032 (0.009)	−0.032 (0.014)	0.040 (0.064)	−0.118 (0.031)	−0.576 (0.118)	0.114 (0.009)
5–9	−0.024 (0.008)	−0.024 (0.014)	0.042 (0.066)	−0.114 (0.023)	−0.407 (0.137)	0.106 (0.008)
10–14	−0.021 (0.010)	−0.021 (0.016)	0.055 (0.083)	−0.137 (0.026)	−0.633 (0.067)	0.082 (0.006)
15–19	−0.020 (0.012)	−0.020 (0.020)	0.071 (0.103)	−0.167 (0.034)	−0.513 (0.070)	0.076 (0.005)
20–24	−0.019 (0.013)	−0.020 (0.022)	0.082 (0.114)	−0.182 (0.041)	−0.554 (0.251)	0.107 (0.017)
25–29	−0.018 (0.013)	−0.019 (0.022)	0.087 (0.112)	−0.179 (0.043)	−0.677 (0.079)	0.061 (0.006)
30–34	−0.017 (0.012)	−0.018 (0.020)	0.087 (0.101)	−0.161 (0.042)	−0.665 (0.061)	0.053 (0.004)
35–39	−0.016 (0.010)	−0.016 (0.017)	0.083 (0.085)	−0.135 (0.039)	−0.728 (0.072)	0.038 (0.003)
40–44	−0.014 (0.008)	−0.014 (0.014)	0.078 (0.067)	−0.106 (0.036)	−0.739 (0.045)	0.037 (0.002)
45–49	−0.012 (0.007)	−0.013 (0.011)	0.073 (0.051)	−0.081 (0.034)	−0.538 (0.074)	0.035 (0.002)
50–54	−0.011 (0.005)	−0.011 (0.009)	0.069 (0.038)	−0.060 (0.033)	−0.472 (0.093)	0.033 (0.002)
55–59	−0.009 (0.004)	−0.009 (0.007)	0.067 (0.029)	−0.045 (0.032)	−0.458 (0.074)	0.031 (0.002)
60–64	−0.008 (0.004)	−0.008 (0.006)	0.065 (0.022)	−0.033 (0.032)	−0.383 (0.091)	0.025 (0.002)
65–69	−0.006 (0.003)	−0.006 (0.005)	0.064 (0.017)	−0.024 (0.032)	−0.559 (0.107)	0.025 (0.002)
70–74	−0.005 (0.003)	−0.005 (0.005)	0.064 (0.014)	−0.017 (0.032)	−0.538 (0.131)	0.021 (0.002)
75–79	−0.004 (0.003)	−0.004 (0.005)	0.065 (0.011)	−0.012 (0.032)	−0.538 (0.079)	0.025 (0.003)
80–84	−0.003 (0.003)	−0.003 (0.005)	0.065 (0.009)	−0.007 (0.032)	−0.481 (0.060)	0.034 (0.004)
85+	−0.002 (0.003)	−0.002 (0.005)	0.066 (0.008)	−0.002 (0.033)	−0.379 (0.117)	0.047 (0.004)
Log-likelihood	4024.39					

The table shows the estimated parameters of the model driven by a two-factor moving average latent process.
This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

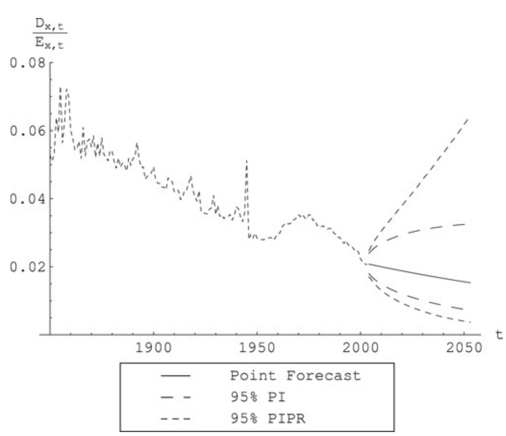

Fig. 4. Prediction age group 65–69 year; one-factor AR. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the one-factor first-order autoregressive model for the age group of 65–69 years, both without (PI) and with (PIPR) taking into account parameter estimation risk.

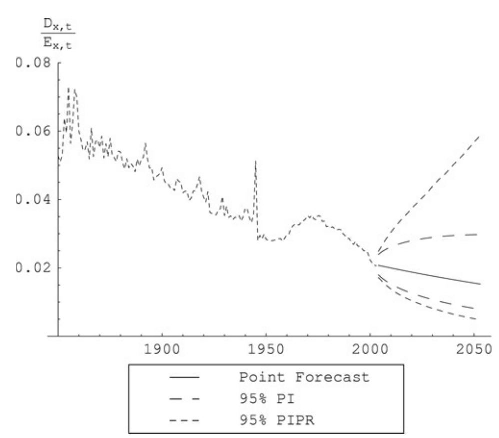

Fig. 5. Prediction age group 65–69 year; one-factor MA. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the one-factor first-order moving average model for the age group of 65–69 years, both without (PI) and with (PIPR) taking into account parameter estimation risk.

use the less accurately, but likely more robustly, estimated $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$. Using in these models also the more accurately estimated $\widehat{A}$ would have resulted in more or less the same point forecasts, but smaller confidence intervals, comparable to the Lee and Carter (1992)-case. However, by using $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ instead of $\widehat{A}$ one might incorporate the model's difficulty in estimating the long run trend. This comes at the cost of a higher inaccuracy.

5. Summary and conclusion

In this paper we extended the often found empirical version of the Lee and Carter (1992) approach, as reformulated by Girosi and King (2005). In this reformulation the log central death rates (or some other way to measure log mortalities) are directly modeled as random walks with drift. These drifts determine the long run forecasts. However, the estimation of these drifts might be rather sensitive to the sample period

Table 7
Model performance

Age group (x)	Cumulative squared deviation 1850–2003				
	LC	1F AR	1F MA	2F AR	2F MA
1–4	3.52	3.62	3.64	3.51	3.62
5–9	3.95	4.18	4.20	4.17	4.24
10–14	4.27	4.43	4.43	4.40	4.44
15–19	4.55	4.96	4.94	4.97	5.02
20–24	7.33	7.40	7.38	7.33	7.26
25–29	6.84	6.87	6.82	6.81	6.74
30–34	5.87	5.95	5.89	5.93	5.83
35–39	4.25	4.29	4.25	4.23	4.19
40–44	3.21	3.11	3.09	3.06	3.04
45–49	2.11	2.14	2.13	2.12	2.09
50–54	1.40	1.47	1.46	1.45	1.40
55–59	1.09	1.14	1.13	1.13	1.09
60–64	0.94	0.98	0.97	0.97	0.94
65–69	0.82	0.87	0.86	0.85	0.82
70–74	0.74	0.77	0.77	0.77	0.74
75–79	0.81	0.85	0.84	0.84	0.81
80–84	0.81	0.85	0.85	0.83	0.81
85+	0.97	1.07	1.06	1.03	0.98

The table compares the in-sample fit of different models: Lee–Carter (LC), single-factor autoregressive (1F AR), single-factor moving average (1F MA), two-factor autoregressive (2F AR) and two-factor moving average (2F MA), based on the cumulative sum of squared deviations of one period ahead in-sample age-specific forecasts.

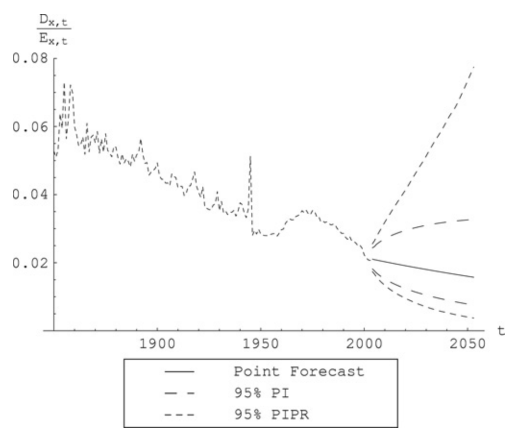

Fig. 6. Prediction age group 65–69 year; two-factor AR. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the two-factor first-order autoregressive model for the age group of 65–69 years, both without (PI) and with (PIPR) taking into account parameter estimation risk.

employed. We extended this approach by allowing for a time-varying trend, depending upon a few underlying latent factors, in order to capture the comovements between the various age groups. We formulated our model in a state-space framework, so that the Kalman filtering technique can be used to estimate the parameters by means of quasi-maximum likelihood.

We illustrated our specification using Dutch mortality data over the period 1850–2003. In particular, we illustrated how our approach might yield a more stable estimation for the long run trend, by incorporating the model's difficulty in estimating the long run trend. When using the whole available sample period, we found comparable estimates for the trend based upon the various approaches, indicating that this sample seems

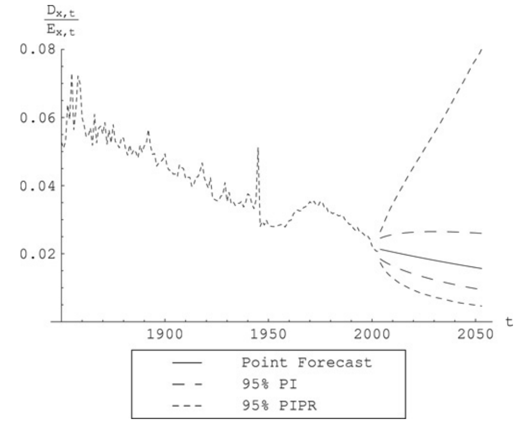

Fig. 7. Prediction age group 65–69 year; two-factor MA. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the two-factor first-order moving average model for the age group of 65–69 years, both without (PI) and with (PIPR) taking into account parameter estimation risk.

to be 'representative' for the long run trend. However, since estimating the long run trend is harder when incorporating the model's difficulty in estimating the long run trend than simply estimating the long run trend by the sample trend, the prediction intervals based on the first estimated long run trends are wider than those obtained by the second approach.

Appendix. Spline interpolation

In this appendix we describe the spline interpolation method that we employ in order to reduce the number of parameters and to avoid localized age-induced anomalies.

(i) First, let $\widetilde{x}_j$ denote the average age for age group j, $j = 1, \ldots, na$, where the age groups are in increasing order. Then the set $\{\widetilde{x}_1, \ldots, \widetilde{x}_{na}\}$ is mapped to [0, 1] in the following way:

$$\{\widetilde{x}_1, \ldots, \widetilde{x}_{na}\} \ni \widetilde{x}_j \rightarrow \overline{x}_j = \frac{\widetilde{x}_j}{\widetilde{x}_{na}} \in [0, 1], \quad j = 1, \ldots, na. \tag{39}$$

(ii) For a we can define the function

$$F^a : \{\overline{x}_1, \ldots, \overline{x}_{na}\} \ni \overline{x}_j \rightarrow a_j \in \mathbb{R}. \tag{40}$$

Similarly, for each factor i in B, we can define the function

$$F_i^B : \{\overline{x}_1, \ldots, \overline{x}_{na}\} \ni \overline{x}_j \rightarrow B_{j,i} \in \mathbb{R}. \tag{41}$$

(iii) The interval [0, 1] is split into three intervals $[0, x^*] \cup [x^*, x^{**}] \cup [x^{**}, 1]$, where $x^* = 20/110$ divides the young age and adult mortality, and $x^{**} = 50/110$ separates the adult and old mortality, since these have different behavior (see Heligman and Pollard (1980)), therefore, sensitivities. The functions F^a and F_i^B are approximated by cubic spline functions. For example, F^a is approximated by $\widetilde{F}^a$, where

$$\begin{aligned} \widetilde{F}^a(x) &= S_l^a(x) \quad \text{if } x \in [0, x^*], \\ &= S_c^a(x) \quad \text{if } x \in [x^*, x^{**}], \\ &= S_r^a(x) \quad \text{if } x \in [x^{**}, 1] \end{aligned} \tag{42}$$

and where

$$S_l^a(x) = l_0^a + l_1^a \times x + l_2^a \times x^2 + l_3^a \times x^3, \tag{43}$$

$$S_c^a(x) = c_0^a + c_1^a \times x + c_2^a \times x^2 + c_3^a \times x^3, \tag{44}$$

$$S_r^a(x) = r_0^a + r_1^a \times x + r_2^a \times x^2 + r_3^a \times x^3. \tag{45}$$

We require this approximation to satisfy the smoothness conditions:

$$S_l^a(x^*) = S_c^a(x^*), \qquad \frac{\partial S_l^a(x^*)}{\partial x} = \frac{\partial S_c^a(x^*)}{\partial x},$$

$$\frac{\partial^2 S_l^a(x^*)}{\partial x^2} = \frac{\partial^2 S_c^a(x^*)}{\partial x^2},$$

$$S_c^a(x^{**}) = S_r^a(x^{**}), \qquad \frac{\partial S_c^a(x^{**})}{\partial x} = \frac{\partial S_r^a(x^{**})}{\partial x} \quad \text{and}$$

$$\frac{\partial^2 S_c^a(x^{**})}{\partial x^2} = \frac{\partial^2 S_r^a(x^{**})}{\partial x^2}.$$

After solving the system of six equations, six parameters, for example, c_1^a, c_2^a, c_3^a, r_1^a, r_2^a, and r_3^a are uniquely determined by l_0^a, l_1^a, l_2^a, l_3^a, c_0^a, and r_0^a. Consequently, using splines with two knots implies the estimation of six parameters for a, and, similarly, six parameters per factor in B.

(iv) Finally, the model can be written in terms of the parameters of the splines, i.e., a and B are defined by

$$a = \begin{pmatrix} \widetilde{F}^a(\overline{x}_1) \\ \vdots \\ \widetilde{F}^a(\overline{x}_{na}) \end{pmatrix}, \quad B = \begin{pmatrix} \widetilde{F}_1^B(\overline{x}_1) & \cdots & \widetilde{F}_{nf}^B(\overline{x}_1) \\ \vdots & \ddots & \vdots \\ \widetilde{F}_1^B(\overline{x}_{na}) & \cdots & \widetilde{F}_{nf}^B(\overline{x}_{na}) \end{pmatrix}.$$

References

Benjamin, B., Pollard, J.H., 1993. The Analysis of Mortality and Other Actuarial Statistics. Institute of Actuaries and Faculty of Actuaries.

Booth, H., Maindonald, J., Smith, L., 2002. Applying Lee–Carter under conditions of variable mortality decline. Population Studies 56, 325–336.

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

Carter, L.R., Prskawetz, A., 2001. Examining structural shifts in mortality using the Lee–Carter method, MPIDR Working Paper WP 2001-007.

Coale, A., Guo, G., 1989. Revised regional model life tables at very low levels of mortality. Population Index 55, 613–643.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Gerber, H.U., 1997. Life Insurance Mathematics. Springer, Zürich.

Girosi, F., King, G., 2005. A reassessment of the Lee–Carter mortality forecasting method, Working Paper, Harvard University.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Heligman, L., Pollard, J.H., 1980. The age pattern of mortality. Journal of the Institute of Actuaries 107, 49–80.

Lee, R.D., 2000. The Lee–Carter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal 4, 80–93.

Lee, R.D., Carter, L.R., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87, 659–671.

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

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

Renshaw, A.E., Haberman, S., 2003b. On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics 32, 379–401.

Renshaw, A.E., Haberman, S., Hatzopoulos, P., 1996. The modelling of recent mortality trends in United Kingdom male assured lives. British Actuarial Journal 2, 449–477.