Available online at www.sciencedirect.com
SCIENCE DIRECT
Insurance: Mathematics and Economics 38 (2006) 556–570
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/ime

A cohort-based extension to the Lee–Carter model for mortality reduction factors

A.E. Renshaw, S. Haberman *

Cass Business School, City University, London, EC1Y 8TZ, UK

Received September 2005; received in revised form December 2005; accepted 9 December 2005

Abstract

The Lee–Carter modelling framework is extended through the introduction of a wider class of generalised, parametric, non-linear models. This permits the modelling and extrapolation of age-specific cohort effects as well as the more familiar age-specific period effects. The choice of error distribution is generalised.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Cohort effects; Mortality reduction factors; Generalised non-linear models; Time series; Mortality projections

1. Introduction

The projection of mortality trends is an important issue in a wide number of areas: for example, planning for social security and health care systems and for the funding of retirement income systems and, in actuarial applications, for the pricing and reserving of annuity portfolios. The Lee–Carter methodology has proved to be an elegant and effective method of forecasting demographic variables including mortality (Lee and Carter, 1992), which are gaining wide acceptance: for example, it is used as a benchmark methodology by the US Bureau of the Census.

Mortality reduction factors (CMI, 1999) are seen as an important means of capturing and projecting historic mortality trends (Renshaw and Haberman, 2003b). They are central to the extrapolation of mortality trends using the Lee–Carter methodology (Lee, 2000) and its Poisson based equivalent (Brouhns et al., 2002a). These methods, as originally constituted, consider the modelling and projection of age–period effects, to the exclusion of possibly noteworthy cohort effects-identified, for example, by the analysis of UK data by Willets (2004). We also recall that the Lee–Carter age–period model does not always fit empirical data well: see, for example, Renshaw and Haberman (2003a). Our objective in the paper is to investigate the feasibility of extending the methodology to the modelling and projection of age–period–cohort effects. The main contribution of this paper is the incorporation of cohort effects into the Lee–Carter methodology.

We begin (Section 2) by presenting the Lee–Carter approach in a wider setting. Given the central role played by the mortality reduction factors in generating mortality projections, we emphasise the targeting of the mortality reduction factor, as opposed to the force of mortality. Thus, while the parametric structure is expanded to allow for age–cohort as well as the familiar age–period effects, the error structure is imposed by specifying the second moment properties

* Corresponding author.
E-mail address: s.haberman@city.ac.uk (S. Haberman).

0167-6687/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2005.12.001

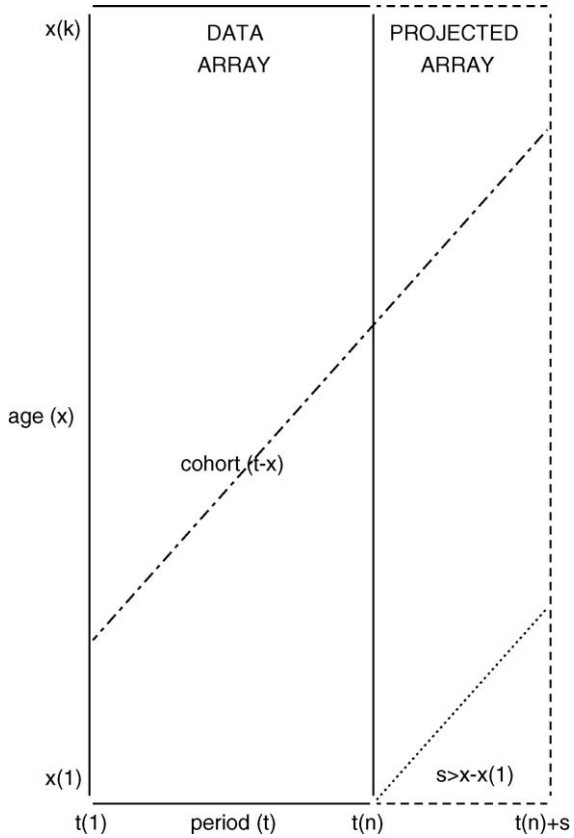

Fig. 1. Observed and projected age–period domains with typical cohort.

of the model. This allows for a range of options for the choice of error distribution including Poisson, both with and without dispersion, as well as Gaussian, as used in the original Lee–Carter approach. Methods of fitting such models are reviewed and expanded. Extrapolation is conducted using the standard approach advocated by Lee and Carter of parametric time series forecasting. The proposed age–period–cohort methodology is applied to the 1961–2000 UK population mortality experience for each gender (Section 3). The patterns in the parameter estimates and the resulting projections are particularly informative and differ appreciably from the results obtained using the standard Lee–Carter approach. A critical appraisal of the methods is presented in Section 4. This includes a study of age specific secular trends (actual and projected) in the rates of change of the logarithms of the force of mortality.

2. Methodology

2.1. Model structure formulation

Let the random variable D_{xt} denote the number of deaths in a population at age x and time t. A rectangular data array (d_{xt}, e_{xt}) is available for analysis where d_{xt} is the actual number of deaths and e_{xt} is the matching exposure to the risk of death. The force of mortality and empirical mortality rates are denoted by μ_{xt} and $\hat{m}_{xt}(= d_{xt}/e_{xt})$, respectively. Cross-classification is by individual calendar year $t \in [t_1, t_n]$ (range n) and by age $x \in [x_1, x_k]$, either grouped into k (ordered) categories, or by individual year (range k), in which case year-of-birth or cohort year $z = t - x \in [t_1 - x_k, t_n - x_1]$ (range $n + k - 1$) is defined. We assume that this is the case throughout. See Fig. 1.

In terms of the force of mortality (as opposed to the central rate of mortality), the Lee–Carter (LC) model structure is

$$\mu_{xt} = \exp(\alpha_x + \beta_x \kappa_t),$$

subject to the usual (non-unique) constraints

$$\sum_{t=t_1}^{t_n}\kappa_t = 0, \qquad \sum_x \beta_x = 1 \tag{1}$$

The LC model structure reduces the dimensionality of the problem by identifying a single time index, which affects the force of mortality at time t at all ages simultaneously. The first constraint under (1) has the effect of centring the κ_t values over the range $t \in [t_1, t_n]$. The structure is designed to capture age–period effects with the α_x terms incorporating the main age effects, averaged over time, and the bilinear terms $\beta_x\kappa_t$ incorporating the age specific period trends (relative to the main age effects). We write

$$\mu_{xt} = \exp(\alpha_x + \log F(x,t))$$

in general, where specifically the mortality reduction factor F

$$\text{LC}: \quad \log F(x,t) = \beta_x\kappa_t$$

is defined under LC modelling. We subsequently adjust the constraints (1), so that

$$\log F(x,t_n) = 0, \quad \forall x$$

when extrapolating mortality rates, as described in Section 2.3.

We now generalise the model structure in order to include age–period–cohort modelling by formulating the mortality reduction factor

$$\text{M}: \quad F(x,t) = \exp(\beta_x^{(0)}\iota_{t-x} + \beta_x^{(1)}\kappa_t),$$

with an extra pair of bilinear terms $\beta_x^{(0)}\iota_{t-x}$ to represent additional cohort effects. Other associated (age–period–cohort) substructures

$$H_0: \quad \beta_x^{(0)} = 1, \beta_x^{(1)} = 1; \quad H_1: \quad \beta_x^{(0)} = 1; \quad H_2: \quad \beta_x^{(1)} = 1$$

for which

$$H_0 \subset H_1; \qquad H_2 \subset \text{M},$$

are also of potential interest when the application of age adjustments to one or both of the main period-effects and cohort-effects terms is not found to be significant. Obviously, LC is the substructure

$$\text{LC}: \quad \beta_x^{(0)} = 0 \quad (\beta_x^{(1)} \equiv \beta_x),$$

and we shall also have occasion to refer to the age–cohort substructure (i.e. with no period effect)

$$\text{AC}: \quad \beta_x^{(1)} = 0 \quad (\beta_x^{(0)} \equiv \beta_x).$$

In formulating these structures, we have partitioned the force of mortality

$$\mu_{xt} = \exp(\alpha_x)F(x,t) \tag{2}$$

into the product of a static life-table incorporating the main age effects α_x and a dynamic parameterised mortality reduction factor F incorporating the age-specific period (κ_t) and cohort (ι_{t-x}) effects.

2.2. Error structure and model fitting

Setting aside fitting by single value decomposition (SVD), which is well documented, it is informative to survey other likelihood fitting methods. Following the general consensus in actuarial modelling, we select the Poisson response model with the response variable equal to the number of deaths. (Direct modelling of the number is very useful for practical applications where, for example, we might need to simulate the future cash flows of an annuity or pension

portfolio.) We allow also for over-dispersion and the allocation of prior weights, which is important in the presence of empty data cells. This is formalised by specifying the first two moments of the responses Y_{xt} where

$$Y_{xt} = D_{xt}, \qquad E(Y_{xt}) = e_{xt}\mu_{xt} = e_{xt}\exp(\alpha_x)F(x,t), \qquad \mathrm{Var}(Y_{xt}) = \phi E(Y_{xt})\left(= \phi\frac{V(E(Y_{xt}))}{w_{xt}}\right) \tag{3}$$

with scale parameter ϕ, variance function $V(E(Y_{xt})) = E(Y_{xt})$ and prior weights $w_{xt} = 1$ (0 if the data cell is empty). Then under the log link, the non-linear predictor η_{xt} is defined as

$$(\log E(Y_{xt}) =)\eta_{xt} = \log e_{xt} + \alpha_x + \log F(x,t).$$

As a possible alternative error structure it is also of interest to note that the original LC Gaussian error structure is re-established on replacing (3) with

$$Y_{xt} = \log\left(\frac{D_{xt}}{e_{xt}}\right), \qquad E(Y_{xt}) = \alpha_x + \log F(x,t), \qquad \mathrm{Var}(Y_{xt}) = \frac{\phi}{w_{xt}}, \tag{4}$$

comprising a free standing scale parameter $\phi(=\sigma^2)$, variance function $V(E(Y_{xt})) = 1$ and prior weights $w_{xt} = 1$. Then, under the identity link, the non-linear predictor is given by

$$(E(Y_{xt}) =)\eta_{xt} = \alpha_x + \beta_x\kappa_t.$$

Given the non-linear nature of the parametric predictors (η_{xt}) we focus on two alternative model fitting procedures: Method A based on an unpublished technical report by Wilmoth (1993), and Method B based on a method of mortality analysis incorporating age–year interaction with applications to medical statistics by James and Segal (1982), which predates Lee and Carter (1992). We present Methods A and B in the context of the Lee–Carter model and then go on to described the fitting procedures for the new models, M and AC, described above.

2.2.1. LC Model

2.2.1.1. Method A. Adapting Wilmoth (1993) and Brouhns et al. (2002a), maximum likelihood estimates are obtained under the original Lee–Carter Gaussian error structure (4) using an iterative process, which can be re-expressed as follows:

$$\begin{array}{c}
\text{Set starting values } \hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t;\ \text{compute } \hat{y}_{xt} \\
\downarrow \\
\text{update } \hat{\alpha}_x;\ \text{compute } \hat{y}_{xt} \\
\text{update } \hat{\kappa}_t,\ \text{adjust s.t. } \displaystyle\sum_{t=t_1}^{t_n}\kappa_t = 0;\ \text{compute } \hat{y}_{xt} \\
\text{update } \hat{\beta}_x;\ \text{compute } \hat{y}_{xt} \\
\text{compute } D(y_{xt},\ \hat{y}_{xt}) \\
\downarrow \\
\text{Repeat the updating cycle; stop when } D(y_{xt},\ \hat{y}_{xt}) \text{ converges}
\end{array}$$

where

$$y_{xt} = \log\hat{m}_{xt}, \qquad \hat{y}_{xt} = \hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t,$$

$$D(y_{xt},\hat{y}_{xt}) = \sum_{x,t}\mathrm{dev}(x,t) = \sum_{x,t}2w_{xt}\int_{\hat{y}_{xt}}^{y_{xt}}\frac{y_{xt}-u}{V(u)}\,\mathrm{d}u = \sum_{x,t}w_{xt}(y_{xt}-\hat{y}_{xt})^2$$

with weights

$$w_{xt} = \begin{cases} 1, & e_{xt} > 0 \\ 0, & e_{xt} = 0 \end{cases}. \tag{5}$$

The updating of a typical parameter θ proceeds according to

$$\text{updated}\,(\theta) = u(\theta) = \theta - \frac{\partial D/\partial\theta}{\partial^2 D/\partial\theta^2}$$

where D is the deviance of the current model. See Table 1 for details. Effective starting values, conforming to the usual Lee–Carter constraints (1) are $\hat{\kappa}_t = 0$, $\hat{\beta}_x = 1/k$, coupled with the SVD estimate

$$\hat{\alpha}_x = \log \prod_{t=t_1}^{t_n} \hat{m}_{xt}^{1/n} \tag{6}$$

so that α_x is estimated by the logarithm of the geometric mean of the empirical mortality rates. The model has $\nu = (k-1)(n-2)$ degrees of freedom. This iterative fitting process generates maximum likelihood estimates under the Poisson error structure (3) (Alho, 2000; Brouhns et al., 2002a; Wilmoth, 1993) on setting

$$y_{xt} = d_{xt}, \qquad \hat{y}_{xt} = \hat{d}_{xt} = e_{xt}\exp(\hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t)$$

$$D(y_{xt}, \hat{y}_{xt}) = \sum_{x,t} \text{dev}(x,t) = \sum_{x,t} 2w_{xt} \int_{\hat{y}_{xt}}^{y_{xt}} \frac{y_{xt} - u}{V(u)}\,\mathrm{d}u = \sum_{x,t} 2w_{xt}\left\{ y_{xt}\log\left(\frac{y_{xt}}{\hat{y}_{xt}}\right) - (y_{xt} - \hat{y}_{xt})\right\}$$

Brouhns et al. (2002a) attribute the iterative method for estimating log-linear models with bilinear terms to Goodman (1979). See Table 1 for details of the parameter updating relationships.

2.2.1.2. Method B. Following James and Segal (1982), we use the iterative procedure:

Set starting values $\hat{\beta}_x$
↓
given $\hat{\beta}_x$, update $\hat{\alpha}_x$, $\hat{\kappa}_t$
given $\hat{\kappa}_t$, update $\hat{\alpha}_x$, $\hat{\beta}_x$
compute $D(y_{xt}, \hat{y}_{xt})$
↓
Repeat the updating cycle; stop when $D(y_{xt}, \hat{y}_{xt})$ converges

Given $\hat{\beta}_x$ or $\hat{\kappa}_t$, updating is by selecting the desired generalised linear model and fitting the predictor, which is linear in the respective remaining parameters. Thus, log-link Poisson responses $y_{xt} = d_{xt}$ with offsets $\log e_{xt}$ are declared in order to generate the same results as the iterative fitting process of Brouhns et al. (2002a). The respective predictors are declared by accessing the model formulae (design matrices), a feature available in GLIM (Francis et al., 1993), for example. In specifying the model formulae, we impose the constraints

$$\kappa_{t_1} = 0, \qquad \sum_x \beta_x = 1,$$

reverting back to the Lee–Carter constraints (1) once convergence is attained.

2.2.2. M Model (H_0, H_1, H_2)

It is well known that age–period–cohort modelling is problematic, since the three factors are constrained by the relationship

$$\text{cohort} = \text{period} - \text{age}.$$

In order to ensure a unique set of parameter estimates, we resort to a two-stage fitting strategy in which α_x is estimated first, typically as in (6) corresponding to the original Lee–Carter SVD approach. Then, the remaining parameters, those of F, may be estimated by suitably adapting Method B by declaring log-link Poisson responses $y_{xt} = d_{xt}$ and the

Table 1
Parameter updating relationships

	Gaussian	Poisson
LC	$u(\hat{\alpha}_x) = \hat{\alpha}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})}{\sum_t w_{xt}}$	$u(\hat{\alpha}_x) = \hat{\alpha}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})}{\sum_t w_{xt}\hat{y}_{xt}}$
	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x}{\sum_x w_{xt}\hat{\beta}_x^2}$	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x}{\sum_x w_{xt}\hat{y}_{xt}\hat{\beta}_x^2}$
	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\kappa}_t}{\sum_t w_{xt}\hat{\kappa}_t^2}$	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\kappa}_t}{\sum_t w_{xt}\hat{y}_{xt}\hat{\kappa}_t^2}$
M	$u(\hat{\iota}_z) = \hat{\iota}_z + \dfrac{\sum_{\substack{x,t \\ t-x=z}} w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x^{(0)}}{\sum_{\substack{x,t \\ t-x=z}} w_{xt}\hat{\beta}_x^{(0)2}}$	$u(\hat{\iota}_z) = \hat{\iota}_z + \dfrac{\sum_{\substack{x,t \\ t-x=z}} w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x^{(0)}}{\sum_{\substack{x,t \\ t-x=z}} w_{xt}\hat{y}_{xt}\hat{\beta}_x^{(0)2}}$
	$u(\hat{\beta}_x^{(0)}) = \hat{\beta}_x^{(0)} + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\iota}_{t-x}}{\sum_t w_{xt}\hat{\iota}_{t-x}^2}$	$u(\hat{\beta}_x^{(0)}) = \hat{\beta}_x^{(0)} + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\iota}_{t-x}}{\sum_t w_{xt}\hat{y}_{xt}\hat{\iota}_{t-x}^2}$
	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x^{(1)}}{\sum_x w_{xt}\hat{\beta}_x^{(1)2}}$	$u(\hat{\kappa}_t) = \hat{\kappa}_t + \dfrac{\sum_x w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x^{(1)}}{\sum_x w_{xt}\hat{y}_{xt}\hat{\beta}_x^{(1)2}}$
	$u(\hat{\beta}_x^{(1)}) = \hat{\beta}_x^{(1)} + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\kappa}_t}{\sum_t w_{xt}\hat{\kappa}_t^2}$	$u(\hat{\beta}_x^{(1)}) = \hat{\beta}_x^{(1)} + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\kappa}_t}{\sum_t w_{xt}\hat{y}_{xt}\hat{\kappa}_t^2}$
AC	$u(\hat{\alpha}_x)$ computed as above	$u(\hat{\alpha}_x)$ computed as above
	$u(\hat{\iota}_z) = \hat{\iota}_z + \dfrac{\sum_{\substack{x,t \\ t-x=z}} w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x}{\sum_{\substack{x,t \\ t-x=z}} w_{xt}\hat{\beta}_x^2}$	$u(\hat{\iota}_z) = \hat{\iota}_z + \dfrac{\sum_{\substack{x,t \\ t-x=z}} w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\beta}_x}{\sum_{\substack{x,t \\ t-x=z}} w_{xt}\hat{y}_{xt}\hat{\beta}_x^2}$
	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\iota}_{t-x}}{\sum_t w_{xt}\hat{\iota}_{t-x}^2}$	$u(\hat{\beta}_x) = \hat{\beta}_x + \dfrac{\sum_t w_{xt}(y_{xt} - \hat{y}_{xt})\hat{\iota}_{t-x}}{\sum_t w_{xt}\hat{y}_{xt}\hat{\iota}_{t-x}^2}$

augmented offsets $\log e_{xt} + \hat{\alpha}_x$ and adapting the design matrices, together with the constraints

$$\sum_x \beta_x^{(0)} = 1, \qquad \sum_x \beta_x^{(1)} = 1 \text{ and either } \iota_{t_1 - x_k} = 0 \text{ (or } \kappa_{t_1} = 0).$$

Obvious simplifications to the design matrices are needed when fitting the associated sub-models H_1 and H_2, while the iterative element in the fitting procedure is redundant when fitting H_0. We note that the model M has $\nu = k(n-3) - 2(n-2)$ degrees of freedom (excluding any provision for the first stage modelling of α_x). Effective starting values are $\beta_x^{(0)} = \beta_x^{(1)} = 1/k$. Fitting is also possible under Method A, once α_x has been estimated, using the extended definitions of $\hat{y}_{xt}$ and adapting the core of the iterative cycle in accordance with the relevant updating relationships (Table 1). Effective starting values are obtained on setting $\beta_x^{(0)} = \beta_x^{(1)} = 1$ and fitting H_0 to generate starting values for ι_z and κ_t.

2.2.2.1. AC Model. Model identification is conveniently achieved by means of the parameter constraints

$$\iota_{t_1 - x_k} = 0, \qquad \sum_x \beta_x = 1.$$

Model fitting is then possible by reformulation of Method A in terms of α_x, β_x and ι_{t-x}. Thus, ι_{t-x} instead of κ_t is updated in the core of the iterative cycle (subject to the adjustment $\iota_{t_1 - x_k} = 0$), using the replacement updating relationships of Table 1. Fitting is also possible using Method B by replacing κ_t with ι_{t-x} and modifying the design matrices accordingly. A possible strategy for generating starting values is to set $\hat{\beta}_x = 1$ and additionally fit the main effects structure $\alpha_x + \iota_{t-x}$ in accordance with the distributional assumptions under Method A. There are $\nu = (k-1)(n-3)$ degrees of freedom in this model.

2.3. Mortality rate projections

Projected mortality rates

$$\dot{m}_{x,t_n+s} = \hat{m}_{x,t_n} \dot{F}(x, t_n + s), \quad s > 0 \tag{7}$$

are computed by alignment with the latest available mortality rates $\hat{m}_{x,t_n}$. Here,

$$\dot{F}(x, t_n + s) = \exp\left\{\hat{\beta}_x^{(0)}(\tilde{\iota}_{t_n - x + s} - \hat{\iota}_{t_n - x}) + \hat{\beta}_x^{(1)}(\dot{\kappa}_{t_n+s} - \hat{\kappa}_{t_n})\right\}, \quad s > 0$$

for which

$$\lim_{s \to 0} \dot{F}(x, t_n + s) = 1$$

is based on the parameter estimates $\hat{\beta}_x^{(i)}$, $\hat{\iota}_z$, $\hat{\kappa}_t^{(i)}$ and the time series forecasts

$$\{\hat{\iota}_z : z \in [t_1 - x_k, t_n - x_1]\} \mapsto \{\dot{\iota}_{t_n - x_1 + s} : s > 0\}$$

$$\{\hat{\kappa}_t : t \in [t_1, t_n]\} \mapsto \{\dot{\kappa}_{t_n+s} : s > 0\}$$

where

$$\tilde{\iota}_{t_n - x + s} = \begin{cases} \hat{\iota}_{t_n - x + s}, & 0 < s \le x - x_1 \\ \dot{\iota}_{t_n - x + s}, & s > x - x_1 \end{cases}.$$

The time series forecasts are typically generated using univariate ARIMA processes. The random walk with drift (or ARIMA(0,1,0) process) features prominently in the published applications of LC. If no provision for alignment with the latest available mortality rates is made (see Eq. (7)), as originally in Lee and Carter (1992), the extrapolated mortality rates decompose multiplicatively as

$$\dot{m}_{x,t_n+s} = \exp(\hat{\alpha}_x + \hat{\beta}_x^{(0)} \hat{\iota}_{t_n - x} + \hat{\beta}_x^{(1)} \dot{\kappa}_{t_n}) \dot{F}(x, t_n + s), \quad s > 0,$$

which has the same functional form as (2), and can be directly compared with (7).

2.4. Discussion

By specifying the second moment distributional properties when defining the model error structure, the choice of distribution is not restricted to the Poisson and Gaussian distributions, and may indeed be expanded by selecting different variance functions. Empirical evidence suggests that, for all practical purposes, maximum likelihood estimates obtained for LC using the iterative fitting processes under the Gaussian error structure (4), are the same as those obtained under fitting by SVD. Unlike modelling by SVD, however, the choice of weights (5) means that estimation can proceed, in the presence of empty data cells, under the Gaussian, Poisson and any other of the viable error settings. Wilmoth (1993) uses weights $w_{xt} = d_{xt}$ in combination with the Gaussian error setting. Empirical studies reveal that this has the effect of bringing the parameter estimates into close agreement with the Poisson response based estimates. When comparing a range of results obtained under both modelling approaches (with identical model structures), we have found that the same number of iterations is required in order to induce convergence. Convergence is slow when fitting M.

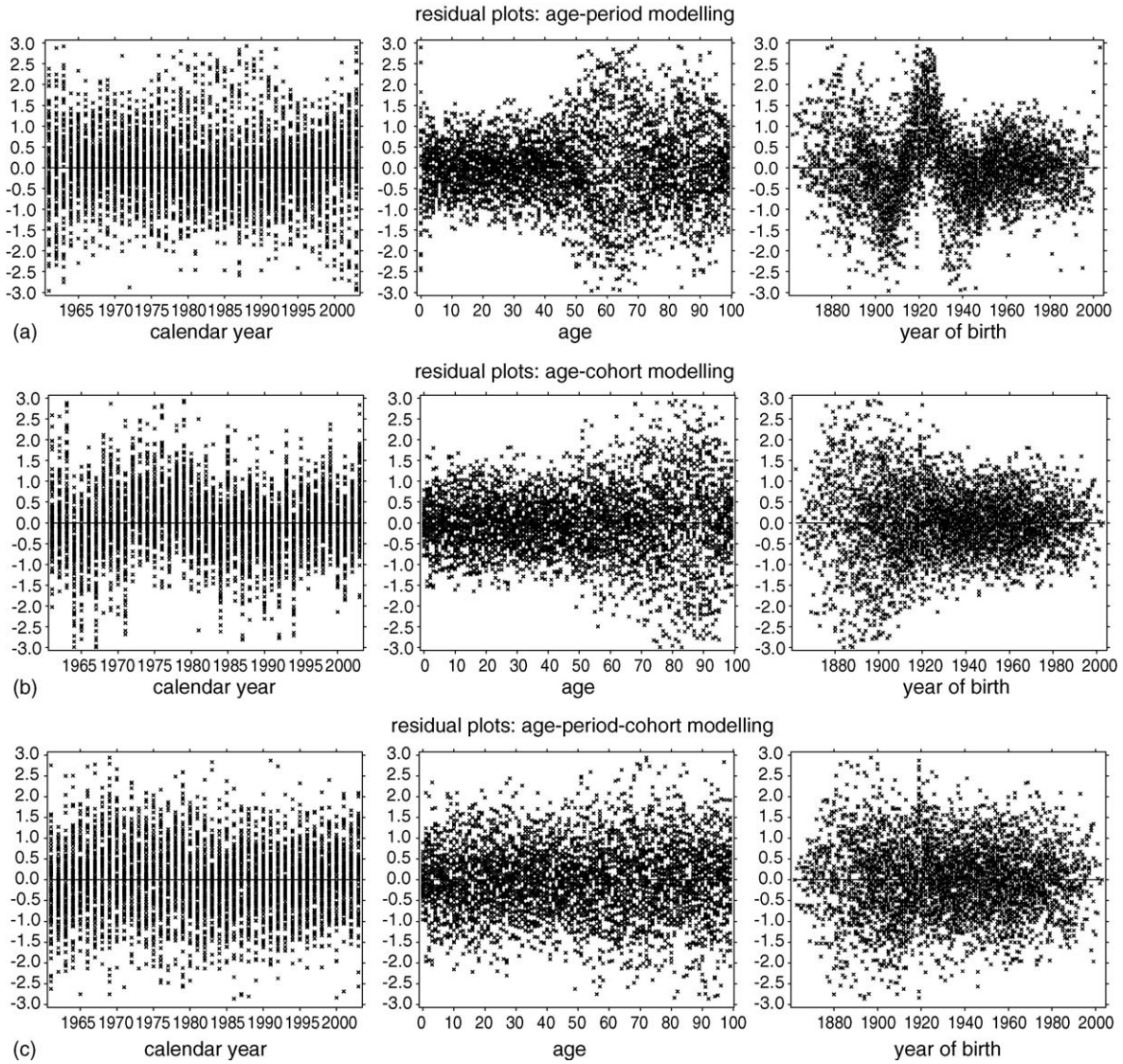

Fig. 2. Female study, residual plots: (a) model LC; (b) model AC; (c) model M.

Diagnostic checks on the fitted model are conducted by plotting the standardised deviance residuals

$$r_{xt} = \text{sign}(y_{xt} - \hat{y}_{xt})\sqrt{\frac{\text{dev}(x,t)}{\hat{\phi}}}, \quad \hat{\phi} = \frac{D(y_{xt}, \hat{y}_{xt})}{\nu}.$$

The sole use of the proportion of the total temporal variance, as measured by the ratio of the first singular value to the sum of singular values under SVD, is not a satisfactory diagnostic indicator, in our experience. However, this index is widely quoted in the demographic literature: see, for example, Tuljapurkar et al. (2000).

The parameters α_x are estimated simultaneously with the parameters of F in both LC and AC modelling. A two-stage estimation process is necessary, however, in which α_x is estimated separately in order to condition on the estimation of F, when fitting M (and its substructures). This two-stage approach can also be applied when fitting LC and AC. For LC modelling, empirical studies show that this has little practical material effect, due to the robust nature of the α_x estimate (6).

3. England and Wales mortality projections

In order to explore the potential of the age–period–cohort model M, we present results for the England and Wales (EW) 1961–2003 mortality experiences for each gender, with cross-classification by individual year of age from 0 to 99. A direct comparison is made with the standard age–period Lee–Carter model LC and the age–cohort model AC, with all models fitted under the Poisson error setting (3).

The implications of the choice of model structure are immediately apparent from the respective residual plots, illustrated for the EW female experience (Fig. 2). Here the distinctive ripple effects in the year-of-birth residual plots under age–period modelling (Fig. 2a, RH frame), signifying a failure to capture cohort effects, is transferred to the calendar-year residual plots under age–cohort modelling (Fig. 2b, LH frame), signifying a reciprocal failure to capture

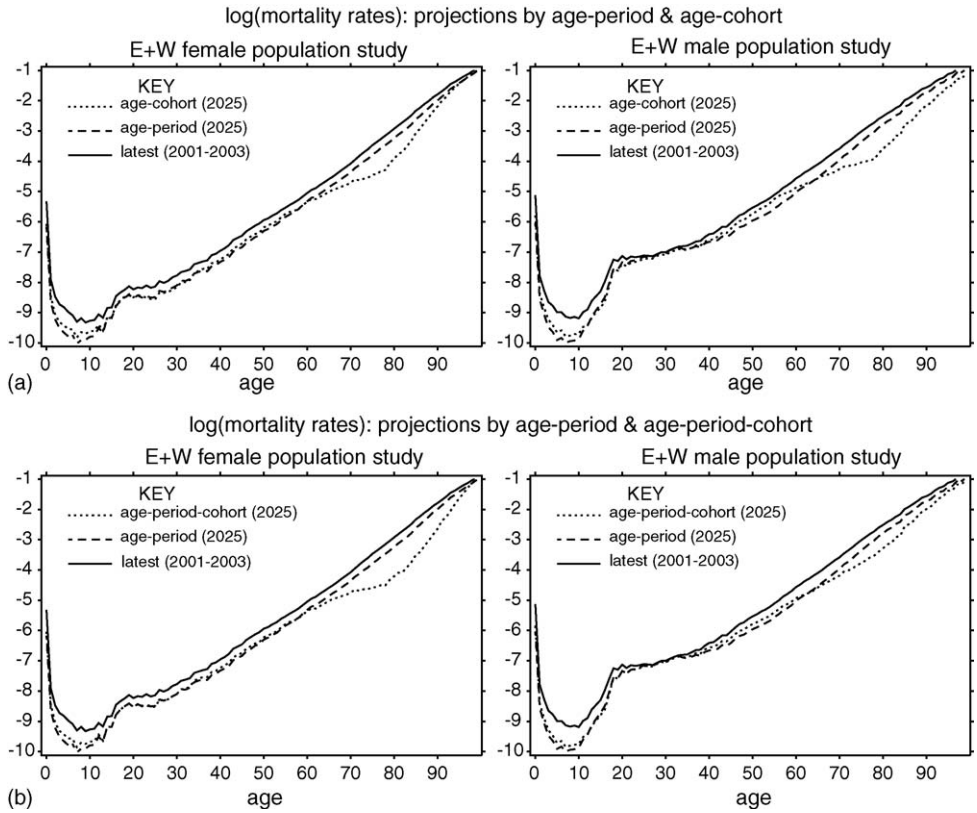

Fig. 3. Latest and projected log μ_{xt} age profiles: (a) LC and AC modelling; (b) LC and M modelling.

period effects. However, the distinctive ripple effects are largely removed under age–period–cohort modelling (Fig. 2c), indicating the relatively successful capture of all three main effects and representing a significant improvement over the fitted Lee–Carter model. Similar patterns are observed in the residual plots for the EW male experience (not reproduced here but the details are available from the authors).

The implications for mortality projections are illustrated (Fig. 3) by plotting the latest available log crude mortality rates (computed by averaging over the last three years) against age for each gender, coupled with the superimposed 2025 projections under the respective age–period (LC) and age–cohort (AC) modelling formulations (Fig. 3a), and under the respective age–period (LC) and age–period–cohort (M) modelling formulations (Fig. 3b). The social and financial implications of these diverse sets of projections are considerable: we attempt to quantify the latter shortly.

Turning first to the parameter estimates for the age–period–cohort modelling approach (Fig. 4), it is informative to compare matching frames between sexes. Thus the main age effect plots ($\hat{\alpha}_x$ versus x) display the familiar characteristics, including 'accident' humps, of static cross-sectional life-tables (on the log scale), with a more pronounced accident

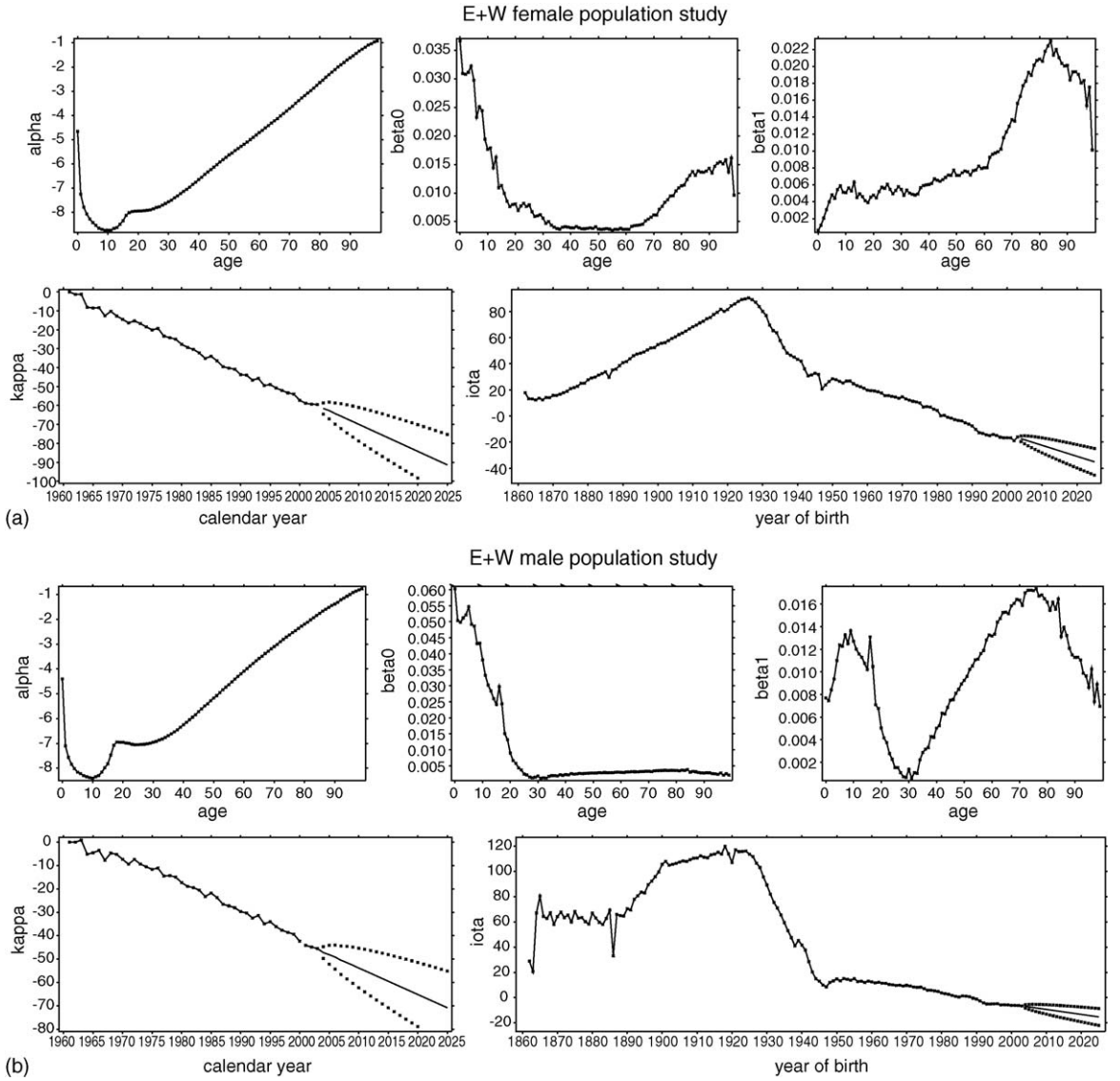

Fig. 4. Parameter estimates, model M: (a) females; (b) males.

hump and heavier mortality for males than for females. Recall that these effects are estimated separately, by averaging crude mortality rates over t for each x, in order to condition for both period and cohort effects.

The main period effect plot ($\hat{\kappa}_t$ versus t) is linear for females but exhibits mild curvature for males, which can be characterised as piece-wise linear with a knot or hinge positioned in the first half of the 1970s. This effect is also present in the separate Lee–Carter analysis of mortality data of the G7 countries (Tuljapurkar et al. (2000)) and has been discussed further for the UK by Renshaw and Haberman (2003a). The forecasts for κ_t are based on the auto-regressive time series

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t, \qquad y_t = \kappa_t - \kappa_{t-1}$$

with $p = 1$ for females, $p = 2$ for males, and with $a_1 = 0$ for both: the equivalent of ARIMA(1,1,0) and ARIMA(2,1,0) modelling, respectively. There are noteworthy differences in the $\hat{\beta}_x^{(1)}$ patterns, which control the rate of decline by period of the age specific rates of mortality in the projections. In particular, the trough in the male $\hat{\beta}_x^{(1)}$ pattern in the 20–40 age range is consistent with similar findings in the reanalysis of the male England and Wales mortality rates (Renshaw and Haberman (2003a)).

The plots of the main cohort effect ($\hat{\iota}_z$ versus $z = t - x$) are particularly revealing. Thus, noteworthy discontinuities occur corresponding to the ending of hostilities in World Wars I and II. While it is possible to identify the first of these with the 1919 influenza epidemic, we are not aware of the likely cause of the second discontinuity. (The 1887 discontinuity can be traced to a set of outliers, and is possibly due to mis-stated exposures for this particular cohort.) The pronounced decline in the $\hat{\iota}_z$ profile in the inter-war years is consistent with the reported rapid mortality improvements experienced by generations born between 1925 and 1945 (for both sexes) and reported by Willets (2004). The apparent

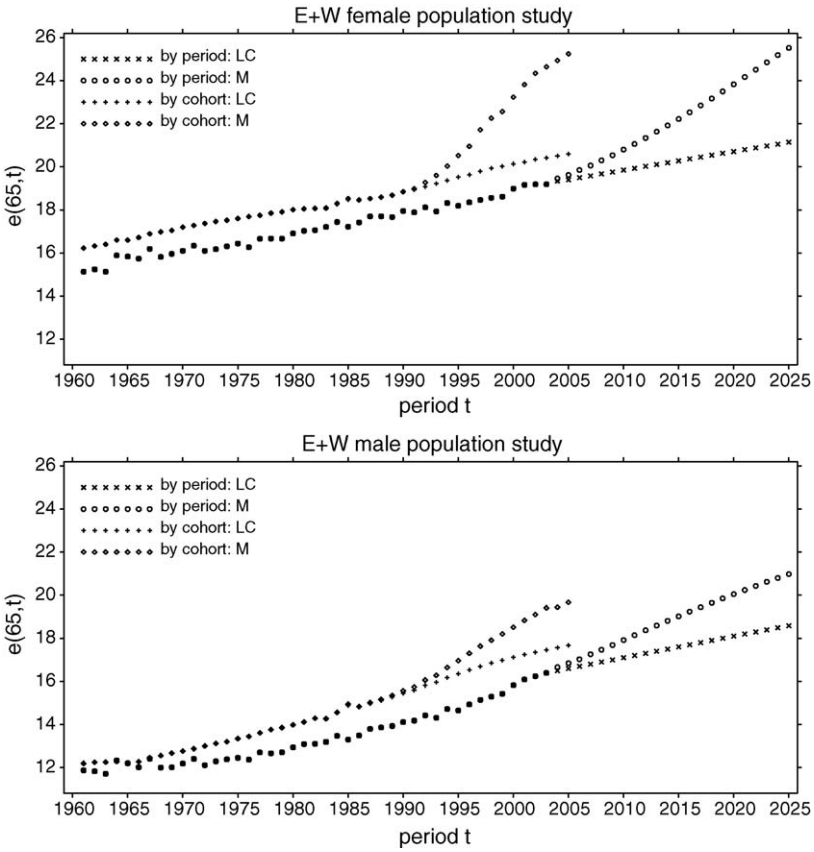

Fig. 5. Life expectancies at age 65 for a range of periods, computed by period and by cohort under age–period (LC) and age–period–cohort (M) modelling.

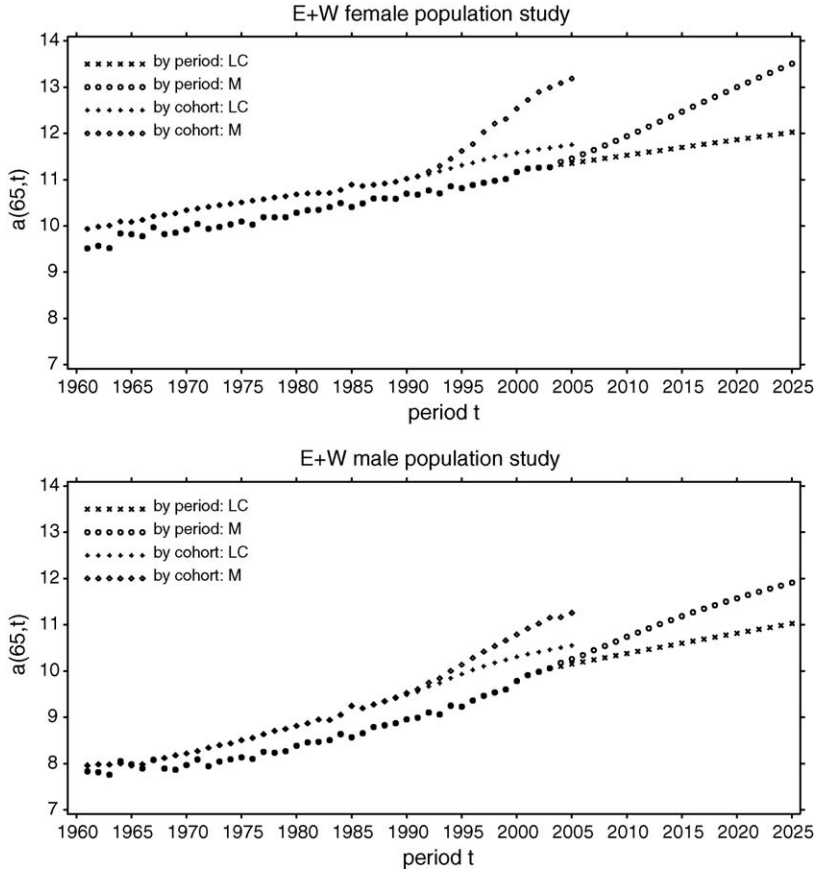

Fig. 6. Annuities, 5% fixed rate interest at age 65 for a range of periods, computed by period and by cohort under age–period (LC) and age–period–cohort (M) modelling.

stable linear trends in the $\hat{\iota}_z$ profiles, present since the late 1940's, form the basis of the depicted time series forecasts, generated using ARIMA(1,1,0) processes. The $\hat{\beta}_x^{(0)}$ patterns, which control the age specific cohort contributions to the mortality projections, are similar, for both sexes, for ages up to 65.

In order to illustrate the impact of such diverse projections under age–period LC and age–period–cohort M modelling, we have calculated complete life expectancies $e_{65}(t)$ at age 65 (Fig. 5) and immediate annuity values $a_{65}(t)$ at age 65 assuming a 5% pa fixed interest rate (Fig. 6) for a range of years t using both the cohort and period method of computing. (We note that the annuity values represent the expected present value of an income of one paid annually in arrears while the individual initially aged 65 remains alive.) Under the cohort method of computing using the formulae

$$e_x(t) = \frac{\sum_{i\geq 0} l_{x+i}(t+i)\left\{1 - \frac{1}{2}q_{x+i}(t+i)\right\}}{l_x(t)}, \qquad a_x(t) = \frac{\sum_{i\geq 1} l_{x+i}(t+i)v^i}{l_x(t)} \tag{8}$$

where

$$q_x(t) \approx 1 - \exp(-\mu_{xt}), \qquad l_{x+1}(t+1) = \{1 - q_x(t)\}l_x(t) \tag{9}$$

with discount factor v, the dynamic aspect of the mortality rates is fully allowed for with the summations proceeding (diagonally) along a cohort. We illustrate values up to the year 2005 calculated by the cohort method and this requires extrapolation up to the year 2040. In contrast, under the period method of calculation, the mortality rates are treated as a sequence of (annual) static life tables, and computing proceeds by suppressing the variation in t in expressions (8) and (9), with (marginal) summation over age ($\geq x$) for each fixed t. We illustrate values up to the year 2025 using this method based solely on empirical mortality rates $\hat{\mu}_{xt} = \hat{m}_{xt}$ in the period up to 2003 and requiring extrapolation

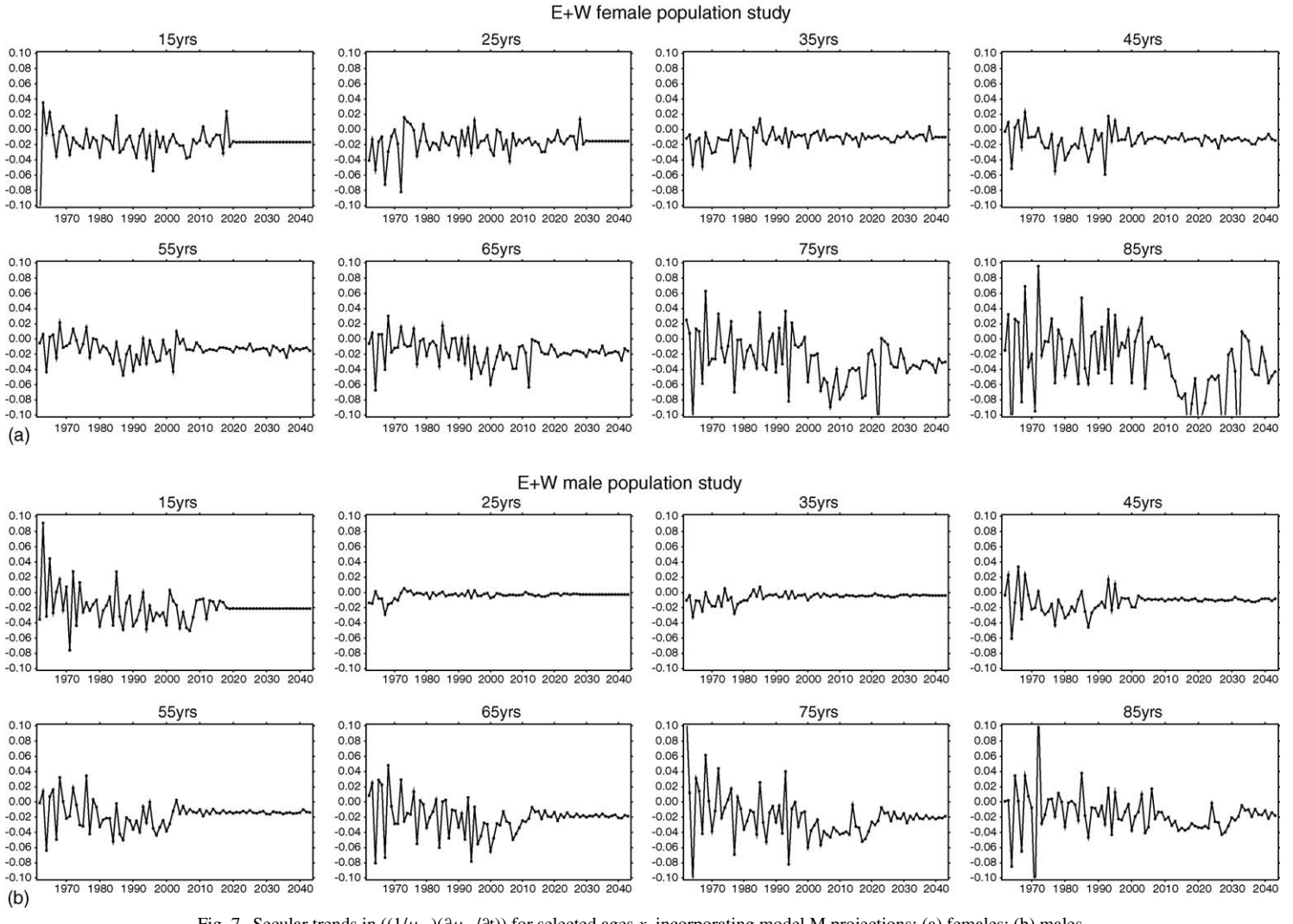

Fig. 7. Secular trends in $((1/\mu_{xt})(\partial\mu_{xt}/\partial t))$ for selected ages x, incorporating model M projections: (a) females; (b) males.

only to the year 2025. On one hand, the period method of computing fails to capture the full dynamic impact of the evolving mortality rates under the modelling assumptions but, on the other hand, generates less uncertainty than the cohort method of computing which requires more lengthy extrapolations and this contributes a source of increasing uncertainty. One means of quantifying this uncertainty is through the adaptation of a Monte Carlo parametric simulation method described by Brouhns et al. (2002b) in the context of LC modelling. This and other methods are currently under investigation. The reserves that insurance companies selling annuities and pension funds would have to hold in order to meet their future contractual liabilities are directly related to terms like $a_{65}(t)$: Booth et al. (2005). The financial implications of the upward trends in Fig. 6 are clear and significant and indicate the burden that increasing longevity may place on such institutions.

4. Critical appraisal

In a reference to the possibility of complicating LC, for example, by adding cohort effects resulting in an improved fit, Lee (2000) comments as follows: 'The question is whether such additional effects would represent fundamental and enduring aspects of mortality patterns such that the forecasts would be improved'. We use this as a basis on which to conduct a critical appraisal of the methodology.

Turning first to the modelling stage and the question of whether the additional structural effects are fundamental aspects of mortality patterns, answers are provided by the analysis of residuals. Thus, on the basis of the residual plots (e.g. Fig. 2), the age–period–cohort model M is shown (Fig. 4) to have captured characteristic and systematic mortality patterns, directly attributable to age, period and cohort effects. Then, if modelling were restricted to LC, distinctive patterns directly attributable to cohort effects would be lost. The question of pattern durability in the future is, however, a matter for conjecture, while established patterns, preferably linear, in the period and cohort parameters are amenable to forecasting.

We turn next to the complication of the extrapolation process through the inclusion of additional age–cohort effects. There is a need to check that the dynamics of the relative contributions from the different sources are maintained in the projections, typically from the age–period and age–cohort effects. In order to do this, we monitor the rate of change of $\log \mu_{xt}$ over t, for each age x, viz.

$$\frac{\partial \log \mu_{xt}}{\partial t} = \frac{\partial \log F(x,t)}{\partial t} = \beta_x^{(0)} \frac{\partial \iota_z}{\partial z} + \sum_{i=1}^{c} \beta_x^{(i)} \frac{\partial \kappa_t^{(i)}}{\partial t} \approx \beta_x^{(0)} \underset{z}{\Delta} \iota_z + \sum_{i=1}^{c} \beta_x^{(i)} \underset{t}{\Delta} \kappa_t^{(i)}$$

where Δ is the differencing operator. A representative selection of such rates of change is depicted in Fig. 7. For both data sets, these plots are consistently linear over the combined data and forecast periods displayed, with the exception of a noteworthy (partially illustrated) trough, which is transmitted by period with increasing age in excess of 65 in the female experience. This departure from linearity for females is associated with the relatively large projected life expectancies and annuities under model M (Figs. 5 and 6).

Acknowledgements

The authors wish to thank the CMI Bureau of the Institute and Faculty of Actuaries for providing financial support for this research, and also Adrian Gallop of the Government Actuary's Department for providing an initial version of the UK population data.

References

Alho, J.M., 2000. Discussion. North American Actuarial Journal 4, 91–93.

Booth, P., Chadburn, R., Haberman, S., James, D., Khorasanee, Z., Plumb, R., Rickayzen, B., 2005. Modern Actuarial Theory and Practice, 2nd ed. CRC Press, Boca Raton.

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

Brouhns, N., Denuit, M., Vermunt, J.K., 2002b. Measuring the longevity risk in mortality projections. Bulletin of the Swiss Association of Actuaries, 105–130.

CMI Committee, 1999. Standard Tables of Mortality Based on the 1991–1994 Experiences. Continuous Mortality Reports, vol. 17. Institute and Faculty of Actuaries, pp. 1–227.

Francis, B., Green, M., Payne, C., 1993. The Glim System: Release 4 Manual. Clarendon, Oxford.
Goodman, L.A., 1979. Simple models for the analysis of association in cross-classifications having ordered categories. Journal of the American Statistics Association 74, 537–552.
James, I.R., Segal, M.R., 1982. On a method of mortality analysis incorporating age–year interaction, with application to prostate cancer mortality. Biometrics 38, 433–443.
Lee, R.D., Carter, L., 1992. Modelling and forecasting the time series of US mortality. Journal of the American Statistics Association 87, 659–671.
Lee, R.D., 2000. The Lee–Carter method of forecasting mortality, with various extensions and applications (with discussion). North American Actuarial Journal 4, 80–93.
Renshaw, A.E., Haberman, S., 2003a. Lee–Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections. Applied Statistics 52, 119–137.
Renshaw, A.E., Haberman, S., 2003b. On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics 32, 379–401.
Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405, 789–792.
Willets, R.C., 2004. The cohort effect: insights and explanations. British Actuarial Journal 10, 833–877.
Wilmoth, J.R., 1993. Computational methods for fitting and extrapolating the Lee–Carter model of mortality change. Technical Report. Department of Demography, University of California, Berkeley.