Identifiability issues of age–period and age–period–cohort models of the Lee–Carter type

Eric Beutner [a,*], Simon Reese [b], Jean-Pierre Urbain [c,1]

[a] Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, Netherlands
[b] Department of Economics Lund University, P.O. Box 7082, SE-22007, Lund, Sweden
[c] Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, Netherlands

ABSTRACT

The predominant way of modelling mortality rates is the Lee–Carter model and its many extensions. The Lee–Carter model and its many extensions use a latent process to forecast. These models are estimated using a two-step procedure that causes an inconsistent view on the latent variable. This paper considers identifiability issues of these models from a perspective that acknowledges the latent variable as a stochastic process from the beginning. We call this perspective the plug-in age–period or plug-in age–period–cohort model. Defining a parameter vector that includes the underlying parameters of this process rather than its realizations, we investigate whether the expected values and covariances of the plug-in Lee–Carter models are identifiable. It will be seen, for example, that even if in both steps of the estimation procedure we have identifiability in a certain sense it does not necessarily carry over to the plug-in models.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Interest in the age-at-death distribution can be traced back to the works of John Graunt and Edmond Halley in 1662 and 1693, respectively; see Hald (2003, Chapters 7 and 9). Deriving analytical expressions for the age-at-death distribution or, which is the same, for the force of mortality goes back to the work of Gompertz in 1825 (or even to de Moivre who assumed a constant force of mortality for his work on annuities). For further analytical expressions for the force of mortality see, for instance, Bowers et al. (1997, Section 3.7). Continuing decrease of mortality rates (and consequently continuing increase of life expectancies) in many developed countries over the last six or seven decades has brought the need of forecasting mortality rates to a leading edge. A prerequisite for extrapolative methods to forecast mortality rates is a model that captures the main features of observed mortality rates. The dominant model of this approach is the Lee–Carter model (cf. Lee and Carter (1992)) and its many variants; for overviews on the original model and on the many extensions that have been proposed one may refer to Booth (2006), Booth and Tickle (2008), Cairns et al. (2008), Cairns et al. (2009), Currie (2016), Haberman and Renshaw (2008) and Haberman and Renshaw (2011), and the references therein.

The basic Lee–Carter model is an age–period model that takes as its starting point a non-linear parametrization of the logarithm of the central forces of mortality. It is given by

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \; x = 0, \ldots, X, t = 1, \ldots, T; \quad (1)$$

cf. first displayed equation in Lee and Carter (1992, Section 3). Here $m_{x,t}$ are the 'observed' central forces of mortality and X is the maximal age (either in the sample or the maximum age of interest). The errors $\epsilon_{x,t}$ are assumed to have mean zero and variance σ_ϵ^2. The $(X + 1)$-dimensional parameter vectors $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_X)$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_X)$ are interpreted as age-specific constants. Because in a first step $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_T)$ is considered to be a T-dimensional parameter vector, the model for the expected values of $\log(m_{x,t})$ defined by (1) is clearly over-parametrized. The solution proposed by Lee and Carter to ensure identifiability of the first moments is to impose the constraints $\sum_{x=0}^{X} \beta_x = 1$ and $\sum_{t=1}^{T} \kappa_t = 0$, cf. first paragraph of Section 3 in Lee and Carter (1992). Under these constraints, called 'ad hoc identification' by Nielsen and Nielsen (2014), $\alpha_x + \beta_x \kappa_t = \tilde{\alpha}_x + \tilde{\beta}_x \tilde{\kappa}_t, x = 0, \ldots, X, t = 1, \ldots, T$ imply that $\alpha_x = \tilde{\alpha}_x, \beta_x = \tilde{\beta}_x, x = 0, \ldots, X$, and $\kappa_t = \tilde{\kappa}_t, t = 1, \ldots, T$. Instead of using what has been called 'ad hoc identification' one can employ the so-called canonical representation which also solves the identifiability issue and provides additional insight into the geometry of the model; see Nielsen and Nielsen (2014, Section 6) and also (Kuang et al., 2008a) for a linear age–period–cohort model. The cohort extension of the age–period model given by Eq. (1) is

* Corresponding author.
E-mail addresses: e.beutner@maastrichtuniversity.nl (E. Beutner), simon.reese@nek.lu.se (S. Reese), j.urbain@maastrichtuniversity.nl (J.-P. Urbain).
[1] Jean-Piere Urbain passed away before the manuscript was submitted.

defined by

$$\log(m_{x,t}) = \alpha_x + \beta_x^{(0)} \iota_{t-x} + \beta_x^{(1)} \kappa_t + \epsilon_{x,t}. \tag{2}$$

Here $\iota = (\iota_{1-X}, \ldots, \iota_T)$ represents cohort effects. This cohort extension was introduced by Renshaw and Haberman (2006). Again this model is over-parametrized; see Section 3.3 for more details on that.

To be able to forecast with models as given by Eqs. (1) and (2) a two step procedure is applied. In a first step the parameters α, β and κ (model (1)) or α, $\beta^{(0)}$, $\beta^{(1)}$, κ and ι (model (2)), respectively, are estimated. In the second step a time series model that allows for forecasting is fitted to the estimated $\hat{\kappa}$-vector or to the estimated $\hat{\kappa}$ and $\hat{\iota}$-vectors. In the following, we will refer to the classical Lee–Carter model and its cohort extension as fully parametric age–period or age–period–cohort Lee–Carter models, respectively. This denomination relates to the treatment of estimated age and cohort effects as factor scores, i.e. estimates of T and $T + X$ dimensional parameter vectors, in step one of the estimation process (see e.g. Rencher (2002, Section 13.6) for the notion of factor scores). Imposing a stochastic model on the estimates of these parameters in step two is conceptually inconsistent and leads to problems when specifying identifiability constraints. For example, it is a priori unclear whether the forecast from the imposed stochastic model depends on the chosen identification scheme for the original parameters α, β and κ (model (1)) or α, $\beta^{(0)}$, $\beta^{(1)}$, κ and ι (model (2)). For model (1) this question has been addressed by Nielsen and Nielsen (2014) who build on Kuang et al. (2008b) where the same question is analysed for an additive age–period–cohort model; for more information on additive age–period–cohort models and an application to a real data set see Kuang et al. (2011). Moreover, as detailed in Section 2, the 'ad hoc identification' constraints lead to implausible constraints on the properties of the stochastic model imposed onto the factor scores. An alternative perspective on the Lee–Carter model is to replace κ, or κ and ι, by time series models from the beginning. We will denote these models as plug-in age–period and age–period–cohort Lee–Carter models respectively. Plug-in age–period Lee–Carter models have so far been considered in Girosi and King (2007), De Jong and Tickle (2006) and Fung et al. (2016). For the latter paper see in particular their remark 2.1.

Recently, Leng and Peng (2016) considered a simplified fully parametric age–period Lee–Carter model and showed that the two step estimation procedure may lead to inconsistent estimators. A pre-requisite for consistency is identifiability. This paper, therefore, considers the interplay between identifiability of fully parametric and plug-in Lee–Carter models. Suppose that we have an identification scheme for a fully parametric Lee–Carter model. Furthermore, suppose that we use an identifiable time series model for κ or identifiable time series models for κ and ι, respectively. Do the plug-in Lee–Carter models inherit identifiability from identifiability of the fully parametric Lee–Carter models and the identified time series models? We show that this needs not be the case. More precisely, we will see that the Lee–Carter model, i.e. (1), under the above mentioned constraints on β and κ combined with a time series model with identified first moments fails to be identifiable in its first moments. Only by adding second moments one can ensure identifiability.

Furthermore, assume that identifiable time series models are plugged in into a non-identified fully parametric Lee–Carter model. Is it possible that the resulting plug-in Lee–Carter model is nevertheless identified? We will see that this possibility can occur. We address these two questions in Section 3 by considering simple but very popular times series models for κ and ι. More precisely, we first look at the age–period model (1) if a random walk is used to model the factor scores for κ. Afterwards we analyse the age–period–cohort Lee–Carter model (2) if two independent random walks are used to model the factor scores for κ and ι. Having

addressed the above questions and having obtained identifiability results if random walks are plugged in we briefly extend our considerations to more complicated time series models because from an applied point it is important that the class of time series models for which plug-in models are identifiable is not too narrow. This will be done in Section 4. Our findings concerning identifiability of plug-in models continue to hold for generalized linear models whose index function is modelled in the fashion of Eqs. (1) and (2).

The rest of the article is organized as follows: In Section 2 we briefly discuss inconsistencies that arise from the two step procedure. Sections 3 and 4 are as described above. These sections are followed by a section where we discuss how forecasts can be obtained with plug-in Lee–Carter models. We conclude with a discussion. All proofs are presented in the appendix.

2. Stochastic process view on the Lee–Carter constraints

If we impose a stochastic model on (κ_t) as done in the statistical analysis of the fully parametric Lee–Carter model, the identifying restriction $\sum_{t=1}^{T} \kappa_t = 0$ becomes a constraint on the possible realizations of the stochastic process (κ_t). As such the constraint does not seem to be sensitive, because it implies inconsistencies in the modelling procedure. An early reference that differentiates sets of constraints depending on whether the factor(s) is/are assumed to be an unobserved random process or unobserved but deterministic is Anderson and Rubin (1956). Here we examine two inconsistencies that arise if we constrain the realizations of the stochastic process (κ_t).

1. *Dynamic view on the constraint:* Suppose that we estimated the model based on data up to and including $\check{T}$ and that we now want to update our estimates based on data up to and including $\check{T} + 1$. If the realization of $(\kappa_1, \ldots, \kappa_{\check{T}})$ fulfils the constraint, then we must have $\kappa_{\check{T}+1} = 0$, because otherwise $\sum_{t=1}^{\check{T}+1} \kappa_t = 0$ is impossible. This is because we cannot change the realization of $(\kappa_1, \ldots, \kappa_{\check{T}})$ which is given to us. This is different from increasing X, because β is part of the modelling process and not given exogenously to us as the realization of a stochastic process. Notice also that the same reasoning applied sequentially to $\check{T} + (k-1)$, $k \geq 2$, would imply $\kappa_{\check{T}+k} = 0$, $k \geq 2$.

2. *Distributional view on the constraint:* Assume, for instance, that the outcome of the second step of the statistical analysis done by Lee and Carter is that (κ_t) follows a random walk with or without drift. Assume additionally that the random walk starts in $c \in \mathbb{R}$, i.e. $\kappa_0 = c$, and that the innovations are normally distributed or more general that the joint distributions of the innovations possess a probability density function with respect to Lebesgue measure then the event $\{\sum_{t=1}^{T} \kappa_t = 0\}$ has probability zero for every $T \geq 1$ regardless of the starting value c, because $\{x \in \mathbb{R}^T | \sum_{i=1}^{T} x_i = 0\}$ is a hyperplane in $\mathbb{R}^T$. Consequently, under these assumptions the probability that the constraint is fulfilled equals zero.

3. Identifiability of plug-in age–period and age–period–cohort Lee–Carter models

3.1. Preliminaries

Throughout and irrespective of whether we consider an age–period or an age–period–cohort model we assume that

$$\mathbb{E}_\theta(\log(m_{x,t})) = f_\theta(x, t), \quad x = 0, \ldots, X, \, t = 1, \ldots, T,$$
$$\mathbb{C}\text{ov}_\theta(\log(m_{x,s}), \log(m_{y,t})) = g_\theta(x, y, s, t),$$
$$x, y \in \{0, \ldots, X\}, \, s, t \in \{1, \ldots, T\},$$

for known functions f and g, and some unknown parameter $\theta \in \Theta$ with the parameter space Θ being a subset of some finite-dimensional space. For some models considered below f_θ depends only on t and not on x. However, in the more general case f_θ depends on x and t. Similar for g_θ. We say that the expected values are identifiable if $f_\theta(x, t) = f_{\tilde{\theta}}(x, t)$ for $x = 0, \ldots, X$, $t = 1, \ldots, T$ implies: $\theta = \tilde{\theta}$. Similarly, we say that the expected values and the covariance structure are identifiable if $f_\theta(x, t) = f_{\tilde{\theta}}(x, t)$ for $x = 0, \ldots, X$, $t = 1, \ldots, T$ and $g_\theta(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$ for $x, y \in \{0, \ldots, X\}$, $s, t \in \{1, \ldots, T\}$, together imply: $\theta = \tilde{\theta}$.

3.2. Identifiability of plug-in age–period models

In this section we analyse plug-in age–period Lee–Carter models if the plug-in process is a random walk, i.e. we assume $\kappa_t = \mu + \kappa_{t-1} + e_t = \mu t + c + \sum_{\ell=1}^{t} e_\ell$, $t \geq 1$, with $\kappa_0 = c$, $c \in \mathbb{R}$, and (e_t) is a sequence of independent and identically distributed random variables with $\mathbb{E}(e_t) = 0$ that is independent of $\epsilon_{x,t}$, $x = 0, \ldots, X$, $t = 1, \ldots, T$. Clearly, there is a one-to-one correspondence between $\mu \in \mathbb{R}$ and $\mathbb{E}(\kappa_t) = c + \mu t$ so that the expected values are identifiable. Looking at the plug-in age–period Lee–Carter model with κ being a random walk we have

$$\log(m_{x,t}) = \alpha_x + \beta_x \mu\, t + \beta_x c + \beta_x \sum_{\ell=1}^{t} e_\ell + \epsilon_{x,t},$$
$$x = 0, \ldots, X, t = 1, \ldots, T. \tag{3}$$

Clearly there is no one-to-one correspondence between $\{(\alpha, \beta, \mu) \in \mathbb{R}^{X+1} \times \mathbb{R}^{X+1} \times \mathbb{R} | \sum_{x=0}^{X} \beta_x = 1\}$ and the expected values of $\log(m_{x,t})$. A counterexample is given by $\mu = 0$ and any two vectors $\beta \neq \tilde{\beta}$ that both fulfil the identification constraint if we put $\tilde{\alpha} = (\alpha_0 + c(\beta_0 - \tilde{\beta}_0), \ldots, \alpha_X + c(\beta_X - \tilde{\beta}_X))$. This reveals that identifiable expected values of a fully parametric age–period model combined with a time series model with identifiable expected values do not lead to identifiable expected values for the plug-in age–period model. Here, the problem can be overcome by excluding $\mu = 0$ or by taking other moments into account. Assume that $\mathbb{E}(e_t^2) = \sigma_e^2$ and consider (3) to be a model with parameter set

$$\Theta_{(0,1,0)} := \left\{ (\alpha, \beta, \mu, \sigma_e^2, \sigma_\epsilon^2) \in \mathbb{R}^{(X+1)} \times \mathbb{R}^{(X+1)} \times \mathbb{R} \times \mathbb{R}_+ \right.$$
$$\left. \times \mathbb{R}_+ | \sum_{x=0}^{X} \beta_x = 1 \right\},$$

where $\mathbb{R}_+ = \{x \in \mathbb{R} | x > 0\}$. For all $\theta \in \Theta$ the expected values and the covariances of (3) are given by

$$\mathbb{E}_\theta(\log(m_{x,t})) = \alpha_x + \beta_x \mu\, t + \beta_x c,$$
$$x = 0, \ldots, X, t = 1, \ldots, T, \tag{4}$$

and

$$\mathbb{C}ov_\theta(\log(m_{x,s}), \log(m_{y,t})) = \beta_x \beta_y \sigma_e^2 \min\{s, t\}$$
$$+ \mathbb{1}_{\{x=y, s=t\}}(x, y, s, t)\sigma_\epsilon^2, \tag{5}$$
$$x, y \in \{0, \ldots, X\}, s, t \in \{1, \ldots, T\},$$

respectively, and we have the following result

Theorem 3.1. *The expected values and the covariance structure of the model given by* (3) *with parameter set* $\Theta_{(0,1,0)}$ *are identifiable in the sense of* Section 3.1 *if* $T \geq 2$.

3.3. Identifiability of plug-in age–period–cohort models

Identifiability of fully parametric age–period–cohort Lee–Carter models has recently been questioned by Hunt and Villegas (2015)

who refer to convergence problems of the maximum likelihood estimator and lack of robustness of the estimates in a number of studies. The identifying restrictions used in these models are usually chosen to be

$$\sum_{x=0}^{X} \beta_x^{(0)} = \sum_{x=0}^{X} \beta_x^{(1)} = 1, \ \kappa_1 = 0, \ \beta_x^{(1)} > 0, \ x = 0, 1, \ldots, X$$

or

$$\sum_{x=0}^{X} \beta_x^{(1)} = \sum_{x=0}^{X} \beta_x^{(0)} = 1, \ \sum_{t=1}^{T} \kappa_t = 0, \ \sum_{h=-X+1}^{T} \iota_h = 0;$$

see, for instance, Haberman and Renshaw (2009, equation (2)) and Yang et al. (2016, Section 3). In fact, as Example 3.2 shows, none of the two sets of constraints above ensures that $(\alpha, \beta^{(1)}, \beta^{(0)}, \kappa, \iota) \neq (\tilde{\alpha}, \tilde{\beta}^{(1)}, \tilde{\beta}^{(0)}, \tilde{\kappa}, \tilde{\iota})$ implies $\alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(0)}\iota_{t-x} \neq \tilde{\alpha}_x + \tilde{\beta}_x^{(1)}\tilde{\kappa}_t + \tilde{\beta}_x^{(0)}\tilde{\iota}_{t-x}$ for at least one pair (x, t).

Example 3.2. Let $X > 2$, $T > 2$. Take $(\alpha, \beta^{(1)}, \kappa)$ such that the constraints on $\beta^{(1)}$ and κ are fulfilled. Put $(\tilde{\alpha}, \tilde{\beta}^{(1)}, \tilde{\kappa}) = (\alpha, \beta^{(1)}, \kappa)$. Next define $\beta^{(0)} = (\beta_0^{(0)}, \ldots, \beta_X^{(0)})$ and $\iota = (\iota_{-X+1}, \ldots, \iota_T)$ by

$$\beta_0^{(0)} = 0.75, \beta_1^{(0)} = 0.25, \beta_x^{(0)} = 0, x = 2, \ldots, X;$$
$$\iota_h = 0 \text{ for } h = -X + 2, \ldots, -1, 1, \ldots, T - 1,$$
$$\iota_{-X+1} = -2, \iota_0 = 1, \iota_T = 1.$$

Then the constraints are clearly fulfilled and

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

Finally define $\tilde{\beta}^{(0)} = (\tilde{\beta}_0^{(0)}, \ldots, \tilde{\beta}_X^{(0)})$ and $\tilde{\iota} = (\tilde{\iota}_{-X+1}, \ldots, \tilde{\iota}_T)$ by

$$\tilde{\beta}_0^{(0)} = 0.5, \tilde{\beta}_1^{(0)} = 0.5, \tilde{\beta}_x^{(0)} = 0, x = 2, \ldots, X;$$
$$\tilde{\iota}_h = 0 \text{ for } h = -X + 2, \ldots, -1, 1, \ldots, T - 1, \tilde{\iota}_{-X+1} = -2,$$
$$\tilde{\iota}_0 = 0.5, \tilde{\iota}_T = 1.5.$$

Again the constraints are fulfilled and $\tilde{\beta}_x^{(0)}\tilde{\iota}_{t-x} = \beta_x^{(0)}\iota_{t-x}$ for all $x = 0, 1, \ldots, X$, $t = 1, \ldots, T$. Apparently, $(\tilde{\beta}^{(0)}, \tilde{\iota}) \neq (\beta^{(0)}, \iota)$ and therefore

$$(\alpha, \beta^{(1)}, \beta^{(0)}, \kappa, \iota) \neq (\tilde{\alpha}, \tilde{\beta}^{(1)}, \tilde{\beta}^{(0)}, \tilde{\kappa}, \tilde{\iota}),$$

but by construction $\alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(0)}\iota_{t-x} = \tilde{\alpha}_x + \tilde{\beta}_x^{(1)}\tilde{\kappa}_t + \tilde{\beta}_x^{(0)}\tilde{\iota}_{t-x}$ for all $x = 0, 1, \ldots, X$, $t = 1, 2, \ldots, T$.

What does the non-identifiability of fully parametric age–period–cohort Lee–Carter models imply for their plug-in counterpart? It turns out that the plug-in model actually is identifiable. Let κ_t and ι_{t-x} be given by

$$\kappa_t = \kappa_0 + t\mu_1 + \sum_{s=1}^{t} e_s^{(1)},$$
$$\iota_{t-x} = \iota_{-X} + (t - x + X)\mu_0 + \sum_{r=1}^{t-x+X} e_{r-X}^{(0)} \tag{6}$$

where $(e_s^{(1)})$ and $(e_t^{(0)})$ are two independent sequences of independent and identically distributed random variables with expected values equal to zero and finite second moments denoted by $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$, respectively; see, for example, Haberman and Renshaw (2008, Section 6.2). The model that results if $(\kappa_1, \ldots, \kappa_T)$ and $(\iota_{-X+1}, \ldots, \iota_T)$ in Eq. (2) are assumed to be given by Eq. (6) is

considered to be a model with parameter set

$$\Theta^{ct}_{(0,1,0)\times(0,1,0)} := \Big\{ (\boldsymbol{\alpha}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}^{(1)}, \mu_0, \mu_1, \sigma^2_{e_1}, \sigma^2_{e_2}, \sigma^2_{\epsilon}) $$

$$\in \mathbb{R}^{X+1} \times \mathbb{R}^{X+1} \times \mathbb{R}^{X+1} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$$

$$\Big| \sum_{x=0}^{X} \beta^{(i)}_x = 1, \ i = 0, 1, \ \boldsymbol{\beta}^{(0)} \neq \boldsymbol{\beta}^{(1)} \Big\}.$$

With $\iota_{-X} = c_0$ and $\kappa_0 = c_1$ the expected values of the model defined by (2) and (6) are given by

$$\mathbb{E}_{\boldsymbol{\theta}}(\log(m_{x,t})) = \alpha_x + \beta^{(0)}_x c_0 + \beta^{(0)}_x \mu_0(t - x + X) \\ + \beta^{(1)}_x c_1 + \beta^{(1)}_x \mu_1 t, \tag{7}$$

and the covariance structure by

$$\mathbb{C}ov_{\boldsymbol{\theta}}(\log(m_{x,s}), \log(m_{y,t}))$$
$$= \beta^{(0)}_x \beta^{(0)}_y \sigma^2_{e_0} \min\{s - x + X, t - y + X\} + \beta^{(1)}_x \beta^{(1)}_y \sigma^2_{e_1} \ \min\{s, t\}$$
$$+ \mathbb{1}_{\{x=y, s=t\}}(x, y, s, t)\sigma^2_{\epsilon}. \tag{8}$$

We have the following result:

Theorem 3.3. *The expected values and the covariance structure of the model defined by* (2) *and* (6) *with parameter set* $\Theta^{ct}_{(0,1,0)\times(0,1,0)}$ *are identifiable in sense of* Section 3.1 *if* $T > X + 2$ *and* $X > 0$.

Remark 3.4. (i) Without $\boldsymbol{\beta}^{(0)} \neq \boldsymbol{\beta}^{(1)}$ identifiability fails. Let $\mu_0 \neq \mu_1$, take $\tilde{\boldsymbol{\beta}}^{(0)} = \tilde{\boldsymbol{\beta}}^{(1)} = \boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}^{(1)}$, $\tilde{\mu}_0 = \mu_1$, $\tilde{\mu}_1 = \mu_0$, $\tilde{\sigma}^2_{e_1} = \sigma^2_{e_1}$, $\tilde{\sigma}^2_{e_2} = \sigma^2_{e_2}$, $\tilde{\sigma}^2_{\epsilon} = \sigma^2_{\epsilon}$, and put $\tilde{\alpha}_x = \alpha_x - (X - x)\tilde{\beta}^{(0)}_x \tilde{\mu}_0 + (X - x)\tilde{\beta}^{(1)}_x \mu_0$.
(ii) Furthermore, $X > 0$ cannot be dropped. Indeed if $X = 0$ the covariance structure (cf. Eq. (8)) reduces to $\mathbb{C}ov(\log(m_{0,s}), \log(m_{0,t})) = (\sigma^2_{e_0} + \sigma^2_{e_1}) \min\{s, t\} + \sigma^2_{\epsilon}$. Here, $\{\tilde{\sigma}^2_{e_0}, \tilde{\sigma}^2_{e_1}, \tilde{\sigma}^2_{\epsilon}\}$ with $\tilde{\sigma}^2_{\epsilon} = \sigma^2_{\epsilon}$, $\tilde{\sigma}^2_{e_0} = \sigma^2_{e_0} + z$ and $\tilde{\sigma}^2_{e_1} = \sigma^2_{e_1} - z$ for any $z \in (-\sigma^2_{e_0}, \sigma^2_{e_1})$ is a valid reparameterization.
(iii) In contrast to this we do not know whether the constraint $T > X+2$ that is employed in the proof of Theorem 3.3 is necessary.

4. Some extensions

Even though mortality rates are predominantly forecasted by fitting a simple random walk model on the estimated latent trend, some studies use other simple autoregressive integrated moving average processes for this purpose. Examples are the autoregressive integrated moving average processes with parameter (1,1,0) (ARIMA(1,1,0)) applied in Cairns et al. (2005) and the autoregressive integrated moving average processes with parameter (0,1,1) (ARIMA(0,1,1)) in Brouhns et al. (2005) as well as Coelho and Nunes (2011) who use both models. Under ARIMA(1,1,0) we have $(1 - \rho L)(\Delta \kappa - \mu) = e_t$. Here L is the lag operator. In this case, as detailed in Appendix A.2 we have that Eq. (1) becomes

$$\log(m_{x,t}) = \alpha_x + \beta_x \mu t + \beta_x c + \beta_x \sum_{\ell=1}^{t} \sum_{k=0}^{\infty} \rho^k e_{\ell-k} + \epsilon_{x,t},$$
$$x = 0, \dots, X, t = 1, \dots, T. \tag{9}$$

We consider (9) to be a model with parameter set

$$\Theta_{(1,1,0)} := \Big\{ (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \rho, \sigma^2_e, \sigma^2_{\epsilon}) \in \mathbb{R}^{X+1} \times \mathbb{R}^{X+1} \times \mathbb{R}$$
$$\times (-1, 1) \times \mathbb{R}_+ \times \mathbb{R}_+ | \sum_{x=0}^{X} \beta_x = 1 \Big\}.$$

We have the following result similar to Theorem 3.1.

Theorem 4.1. *The expected values and the covariance structure of the model given by* (9) *with parameter set* $\Theta_{(1,1,0)}$ *are identifiable in the sense of* Section 3.1 *if* $T \geq 4$.

Lastly, assuming that the latent factor follows an ARIMA(0,1,1) process amounts to letting $\Delta \kappa_t - \mu = e_t + \phi e_{t-1}$. Consequently, (1) becomes

$$\log(m_{x,t}) = \alpha_x + \beta_x \mu t + \beta_x c + \beta_x \sum_{\ell=1}^{t}(e_\ell + \phi e_{\ell-1}) + \epsilon_{x,t},$$
$$x = 0, \dots, X, t = 1, \dots, T. \tag{10}$$

We consider (10) to be a model with parameter set

$$\Theta_{(0,1,1)} := \Big\{ (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \phi, \sigma^2_e, \sigma^2_{\epsilon}) \in \mathbb{R}^{X+1} \times \mathbb{R}^{X+1} \times \mathbb{R}$$
$$\times (-1, 1) \times \mathbb{R}_+ \times \mathbb{R}_+ | \sum_{x=0}^{X} \beta_x = 1 \Big\}.$$

We have the following result:

Theorem 4.2. *The expected values and the covariance structure of the model given by* (10) *with parameter set* $\Theta_{(0,1,1)}$ *are identifiable in the sense of* Section 3.1 *if* $T \geq 2$.

The proofs of Theorems 4.1 and 4.2 are both given in the appendix.

5. Forecasting with plug-in Lee–Carter models

In this section we consider the plug-in Lee–Carter model, i.e. Eq. (1) with (κ_t) as specified in the second line of Section 3.2 and parameter set $\Theta_{(0,1,0)}$, to exemplify how forecasts can be obtained with plug-in models. To do so we shall rewrite this model as a state space model. Representing mortality models as state space models has recently been considered by Fung et al. (2016). Among other models they also consider the model given by (1) for Gaussian error terms that are allowed to be heteroscedastic.

To rewrite the plug-in Lee–Carter model as state space model define for $t \geq 1$

$$\log(\mathbf{m}_t) = \begin{pmatrix} \log(m_{0,t}) \\ \vdots \\ \log(m_{X,t}) \end{pmatrix}, \ \mathbf{Z}_t = \begin{pmatrix} \kappa_t \\ 1 \end{pmatrix},$$

$$F = \begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}, \ G = \begin{pmatrix} \beta_0 & \alpha_0 \\ \vdots & \vdots \\ \beta_X & \alpha_X \end{pmatrix} \tag{11}$$

and

$$\mathbf{e}_t = \begin{pmatrix} e_t \\ 0 \end{pmatrix}, \ \boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{0,t} \\ \vdots \\ \epsilon_{X,t} \end{pmatrix}.$$

Then we have for $t \geq 1$ that

$$\log(\mathbf{m}_t) = G\mathbf{Z}_t + \boldsymbol{\epsilon}_t, \ \text{and} \ \mathbf{Z}_{t+1} = F\mathbf{Z}_t + \mathbf{e}_{t+1}. \tag{12}$$

We refer to the system in Eq. (12) as the state space representation of the plug-in Lee–Carter model because it fulfils the definition of a state space model; for the latter see, for instance, Brockwell and Davis (2006, Section 12.1). If the parameter vector $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \sigma^2_e, \sigma^2_{\epsilon})$ were known we could use the Kalman filter to obtain estimates or, which seems to be a more appropriate term, reconstructions of the unobserved $\kappa_1, \dots, \kappa_t$ from the observed $\log(\mathbf{m}_\ell)$, $1 \leq \ell \leq t$. Furthermore, and more importantly for forecasting mortality rates, the Kalman filter can also be used to have predictions for the unobserved κ_{t+1} from the observed $\log(\mathbf{m}_\ell)$, $1 \leq \ell \leq t$; for a general introduction to the Kalman filter see, for instance, Brockwell and Davis (2006, Section 12.2). However, if

$(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \sigma_e^2, \sigma_\epsilon^2)$ is unknown we first need estimates of the unknown parameters. One way to obtain estimates is to use Eqs. (4) and (5). These equations suggest regressing $\log(m_{x,t})$ on $\alpha_x + \beta_x \mu\, t$, $0 \leq x \leq X$, $1 \leq t \leq T$, and $\log(m_{x,s}) \log(m_{y,t})$ on $\beta_x \beta_y \sigma_e^2 \min\{s, t\} + \mathbb{1}_{\{x=y,s=t\}}(x, y, s, t)\sigma_\epsilon^2 + (\alpha_x + \beta_x \mu\, s)(\alpha_y + \beta_y \mu\, t)$, $x, y \in \{0, \ldots, X\}$, $s, t \in \{1, \ldots, T\}$, i.e. we minimize

$$\sum_{x=0}^{X} \sum_{t=1}^{T} t^{-1}\big(\log(m_{x,t}) - (\alpha_x + \beta_x \mu\, t)\big)^2 \tag{13}$$

$$+ \sum_{x=0}^{X} \sum_{s=1}^{T} \sum_{y=1}^{X} \sum_{t=1}^{T} (\min\{s, t\})^{-1}\Big(\log(m_{x,s}) \log(m_{y,t}) \tag{14}$$

$$- [\beta_x \beta_y \sigma_e^2 \min\{s, t\} + \mathbb{1}_{\{x=y,s=t\}}(x, y, s, t)\sigma_\epsilon^2$$

$$+ (\alpha_x + \beta_x \mu\, s)(\alpha_y + \beta_y \mu\, t)]\Big)^2 \tag{15}$$

over the set $\Theta_{(0,1,0)}$. We denote the solution to this minimization problem by $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\epsilon^2)$. Of course, from identifiability of the first two moments only we cannot conclude that this estimation method leads to consistent estimators. Notice also that, for instance, the $[\log(m_{x,t}) - (\alpha_x + \beta_x \mu\, t)]$'s are not identically distributed nor independent, and that the variance of these terms increases as a function of t. It might therefore be that other weights in (13) or estimating the parameters in first differences or other variance stabilizing transformations lead to estimators with better statistical properties. However, here the goal is only to exemplify how one can forecast with plug-in models. Notice that we put $c = 0$ in (4) and that in the above theorems c was considered to be a real number not a parameter. Otherwise we estimate $\alpha_x + \beta_x c$. We now continue to explain how forecasts for mortality rates can be obtained using the Kalman filter. To do so we make use of the estimates $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mu}, \hat{\sigma}_e^2$ and $\hat{\sigma}_\epsilon^2$ that will replace the true but unknown values in the Kalman filter. To initialize the Kalman filter we must provide a starting value $\hat{\mathbf{Z}}_1$ for the latent $\mathbf{Z}_1$. This can, for instance, be done by taking the unconditional expectation of $\mathbf{Z}_1$ and replacing the unknown parameters by their estimates. We then arrive at

$$\hat{\mathbf{Z}}_1 = \begin{pmatrix} \hat{\mu} \\ 1 \end{pmatrix}.$$

We further need the following quantities to initialize the Kalman filter

$$\Psi_1 = \mathbb{E}[\hat{\mathbf{Z}}_1 \hat{\mathbf{Z}}_1'] = \begin{pmatrix} \hat{\mu}^2 & \hat{\mu} \\ \hat{\mu} & 1 \end{pmatrix}, \quad \Pi_1 = \mathbb{E}[\mathbf{Z}_1 \mathbf{Z}_1'] = \begin{pmatrix} \hat{\mu}^2 + \hat{\sigma}_e^2 & \hat{\mu} \\ \hat{\mu} & 1 \end{pmatrix},$$

and $\Phi_1 = \Pi_1 - \Psi_1 = \begin{pmatrix} \hat{\sigma}_e^2 & 0 \\ 0 & 0 \end{pmatrix}$,

where A' stands for the transpose of a matrix A and the expectation was taken w.r.t. e_1. Here, as before, we replaced the unknown parameters in $\mathbb{E}[\mathbf{Z}_1 \mathbf{Z}_1']$ by their estimates. Based on the observed values $\log(\mathbf{m}_1), \ldots, \log(\mathbf{m}_T)$ we can now predict $\mathbf{Z}_{T+1}$ using the initial values and the following recursion for $t = 1, \ldots, T$

$$\Delta_t = \hat{G}\Phi_t \hat{G}' + \hat{R}, \ \Xi_t = \hat{F}\Phi_t \hat{G}',$$

$$\hat{\mathbf{Z}}_{t+1} = \hat{F}\hat{\mathbf{Z}}_t + \Xi_t \Delta_t^{-1}(\log(\mathbf{m}_t) - \hat{G}\hat{\mathbf{Z}}_t),$$

$$\Pi_{t+1} = \hat{F}\Pi_t \hat{F}', \ \Psi_{t+1} = \hat{F}\Psi_t \hat{F}' + \Xi_t \Delta_t^{-1}\Xi_t', \ \Phi_{t+1} = \Pi_{t+1} - \Psi_{t+1},$$

where Δ_t^{-1} denotes a (generalized) inverse, $\hat{F}$ and $\hat{G}$ equal F and G (see Eq. (11) for the definition of F and G) with μ replaced by $\hat{\mu}$, and β_x and α_x, $x = 0, \ldots, X$, replaced by $\hat{\beta}_x$ and $\hat{\alpha}_x$, respectively. Furthermore, the $(X+1) \times (X+1)$ matrix $\hat{R}$ is given by $\hat{\sigma}_\epsilon^2\, \mathbb{I}_{(X+1) \times (X+1)}$ with $\mathbb{I}_{(X+1) \times (X+1)}$ denoting the $(X+1) \times (X+1)$ identity matrix. It is worth mentioning that the matrices are such that the second row of $\hat{\mathbf{Z}}_{t+1}$ remains equal to 1. Now from the general formula for predictions of the observable of a state space system, see, for

instance, again (Brockwell and Davis, 2006, Section 12.2) we find that

$$\widehat{\log(\mathbf{m}_{t+1})} = G\hat{\mathbf{Z}}_{t+1}, \ \text{and} \ \ \widehat{\log(\mathbf{m}_{t+h})} = GF^{h-1}\hat{\mathbf{Z}}_{t+1}, \ h = 2, 3, \ldots.$$

It is worth mentioning that F^h, $h = 1, 2, \ldots$ equals

$$\begin{pmatrix} 1 & h\,\mu \\ 0 & 1 \end{pmatrix}.$$

Hence, denoting the prediction of $\hat{\mathbf{Z}}_{t+1}$ obtained by using the Kalman filter by

$$\begin{pmatrix} \hat{\kappa}_{t+1} \\ 1 \end{pmatrix}$$

we see that the h, $h \geq 1$, step ahead forecast $\widehat{\log(m_{x,t+h})}$ of $\log(m_{x,t+h})$ becomes

$$\widehat{\log(m_{x,t+h})} = \hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_{t+1} + \hat{\beta}_x \hat{\mu}\,(h - 1).$$

Comparing this to the Lee–Carter forecast it is worth noting that the structure is very similar. Yet, the latent and stochastic nature of (κ_t) is accounted for by using the Kalman filter. Moreover, using the observed $\log(\mathbf{m}_t)$, $t = 1, \ldots, T$, we can predict the latent κ_{t+1} and therefore only have to iterate $(h - 1)$ times instead of h times the prediction by the expected value.

6. Discussion

Extensions of the seminal model of Lee and Carter (1992) are by now so plentiful that they constitute an own branch of the literature on mortality modelling. However, despite the large number of contributions that focus on improving the performance of the standard model, research on the properties of the pervasive features of all these models is still scarce. This article considers identifiability of age–period and age–period–cohort Lee–Carter models and their extensions from a perspective that directly acknowledges the stochastic nature of the latent trend in mortality. The relevance of viewing models from this perspective, resulting in what we call the plug-in Lee–Carter model, is emphasized by logical inconsistencies that arise from the standard practice of treating the latent trend in mortality as a parameter vector. We find that the predominant set of identifying restrictions need not ensure identifiability of plug-in age–period Lee–Carter models. On the contrary, age–period–cohort plug-in Lee–Carter models may be identified under restrictions that are insufficient to ensure identifiability of their equivalent under the classical perspective. Still, despite the findings provided in this paper and in other recent work with a similar focus as for example by Nielsen and Nielsen (2014), Leng and Peng (2016) and Fung et al. (2016), much work remains to be done on the theoretical properties of the Lee–Carter model and its extensions. Investigating these is of utmost importance for further contributions on estimation of the model and the conclusions that can be drawn from its parameter values.

Acknowledgements

Appendix

This appendix contains the proofs of the theorems given above. Regarding the proofs for age–period models it is worth mentioning that a technique different from ours could be used: First demean these models, second apply results from Heaton and Solo (2004)

to identify the covariance structure up to some indeterminacies, third use the results of the second step and the constraints on $\boldsymbol{\beta}$ to identify the expected values and to overcome the mentioned indeterminacies. We do not follow this approach as it cannot be applied to age–period–cohort models and it would not simplify the proofs. Moreover, Heaton and Solo (2004) consider the covariance structure to be identifiable if $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$, $x, y \in \{0, \ldots, X\}$, $s, t \in \{1, 2, \ldots\}$ implies $\theta = \tilde{\theta}$ which does not allow to directly determine a finite T as it is done in Theorems 3.1, 4.1 and 4.2.

A.1. Proofs of Theorems 3.1 and 3.3

Proof of Theorem 3.1. Let $\theta, \theta \in \Theta$ and assume $f_{\theta}(x, t) = f_{\tilde{\theta}}(x, t)$ and $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$. For $x = y$ and $t = s$ using expression (5) the latter implies $\beta_x^2 \sigma_e^2 t + \sigma_\epsilon^2 = \tilde{\beta}_x^2 \tilde{\sigma}_e^2 t + \tilde{\sigma}_\epsilon^2$. Because both are affine-linear functions and $T \geq 2$ we must have $\beta_x^2 \sigma_e^2 = \tilde{\beta}_x^2 \tilde{\sigma}_e^2$ and $\sigma_\epsilon^2 = \tilde{\sigma}_\epsilon^2$. For $X = 0$ we obtain $\sigma_e^2 = \tilde{\sigma}_e^2$, because then $\beta_0 = \tilde{\beta}_0 = 1$. Furthermore, for $X = 0$ expression (4) yields $(\alpha_0 + c) + \mu t = (\tilde{\alpha}_0 + c) + \tilde{\mu} t$ which implies $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$ and $\mu = \tilde{\mu}$, because $T \geq 2$.

For $X \geq 1$, it is enough to restrict attention to those β_x that are unequal 0, because if $\beta_x = 0$ we must also have $\tilde{\beta}_x = 0$, otherwise $\beta_x^2 \sigma_e^2 = \tilde{\beta}_x^2 \tilde{\sigma}_e^2$ is impossible since $\sigma_e^2 > 0$ and $\tilde{\sigma}_e^2 > 0$. Put $\check{X} := \{x \in \{0, \ldots, X\} | \beta_x \neq 0\}$. For $t = T$ and $x \neq y$ expression (5) implies that $\beta_x \beta_y \sigma_e^2 s = \tilde{\beta}_x \tilde{\beta}_y \tilde{\sigma}_e^2 s$ for $x, y \in \check{X}$ and $s \in \{1, \ldots, T\}$. For x and y fixed this is a linear function in s, hence: $\beta_x \beta_y \sigma_e^2 = \tilde{\beta}_x \tilde{\beta}_y \tilde{\sigma}_e^2$, $x, y \in \check{X}$. This can only hold if either $\text{sgn}(\beta_x) = \text{sgn}(\tilde{\beta}_x)$, $\forall x \in \check{X}$, or if $\text{sgn}(\beta_x) = -\text{sgn}(\tilde{\beta}_x)$, $\forall x \in \check{X}$, where sgn denotes the signum function. In the latter case we have $-K\tilde{\beta}_x = \beta_x$ for the constant $K = \sqrt{\tilde{\sigma}_e^2 / \sigma_e^2}$, which leads to a contradiction since $1 = \sum_{x=0}^{X} \beta_x = -K \sum_{x=0}^{X} \tilde{\beta}_x = -K$ is impossible as $K > 0$. The same reasoning shows that if $\text{sgn}(\beta_x) = \text{sgn}(\tilde{\beta}_x)$, $\forall x \in \check{X}$, we must have $K = 1$ which implies $\beta_x = \tilde{\beta}_x$ and $\sigma_e^2 = \tilde{\sigma}_e^2$. Using $\beta_x = \tilde{\beta}_x$ expression (4) implies $(\alpha_x + c\beta_x) + \beta_x \mu t = (\tilde{\alpha}_x + c\tilde{\beta}_x) + \tilde{\beta}_x \tilde{\mu} t$ which leads to $\alpha_x = \tilde{\alpha}_x$ and $\mu = \tilde{\mu}$ as $T \geq 2$ and $\beta_x = \tilde{\beta}_x \neq 0$ for at least one $x \in \{0, \ldots, X\}$. $\square$

Proof of Theorem 3.3. Let $\theta, \tilde{\theta} \in \Theta$ and assume $f_{\theta}(x, t) = f_{\tilde{\theta}}(x, t)$ and $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$. For $x = y, s = t$ expression (8) leads to

$$((\beta_x^{(0)})^2 \sigma_{e_0}^2 + (\beta_x^{(1)})^2 \sigma_{e_1}^2) t - (\beta_x^{(0)})^2 \sigma_{e_0}^2 (x - X) + \sigma_\epsilon^2$$
$$= ((\tilde{\beta}_x^{(0)})^2 \tilde{\sigma}_{e_0}^2 + (\tilde{\beta}_x^{(1)})^2 \tilde{\sigma}_{e_1}^2) t - (\tilde{\beta}_x^{(0)})^2 \tilde{\sigma}_{e_0}^2 (x - X) + \tilde{\sigma}_\epsilon^2.$$

Because for x fixed both functions are affine linear in t and by assumption $T > 1$, we obtain by considering the case $x = X$ that $\sigma_\epsilon^2 = \tilde{\sigma}_\epsilon^2$. Then by the same argument we get $(\beta_x^{(0)})^2 \sigma_{e_0}^2 = (\tilde{\beta}_x^{(0)})^2 \tilde{\sigma}_{e_0}^2$ for $x = 0, \ldots, X - 1$. As in the proof of Theorem 3.1 we obtain $|\tilde{\beta}_x^{(0)}| = c|\beta_x^{(0)}|$, $x = 0, \ldots, X - 1$. Now taking $x = X$, $y = X - 1$ and $s = t + 2$ the right-hand side of (8) equals

$$\beta_X^{(0)} \beta_{X-1}^{(0)} \sigma_{e_0}^2 (t + 1) + \beta_X^{(1)} \beta_{X-1}^{(1)} \sigma_{e_1}^2 t$$
$$= \beta_X^{(0)} \beta_{X-1}^{(0)} \sigma_{e_0}^2 + (\beta_X^{(1)} \beta_{X-1}^{(1)} \sigma_{e_1}^2 + \beta_X^{(0)} \beta_{X-1}^{(0)} \sigma_{e_0}^2) t. \quad (16)$$

Because by assumption $T \geq 4$ we have $s \leq T$ for at least $t = 1$ and $t = 2$ so that (16) holds for at least two different values of t. Then in view of $|\beta_{X-1}^{(0)} \sqrt{\sigma_{e_0}^2}| = |\tilde{\beta}_{X-1}^{(0)} \sqrt{\tilde{\sigma}_{e_0}^2}|$ and our assumption $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$ it implies $|\beta_X^{(0)} \sqrt{\sigma_{e_0}^2}| = |\tilde{\beta}_X^{(0)} \sqrt{\tilde{\sigma}_{e_0}^2}|$. Hence $\beta_X^{(0)}$ and $\tilde{\beta}_X^{(0)}$ differ by the same multiplicative constant c as $\beta_x^{(0)}$ and $\tilde{\beta}_x^{(0)}$, $x = 0, \ldots, X - 1$, do. Now for $k = 1, \ldots, X$ taking

$x = X, y = X - k$ and $s = t + k + 1$ the right-hand side of (8) equals

$$\beta_X^{(0)} \beta_{X-k}^{(0)} \sigma_{e_0}^2 (t + k) + \beta_X^{(1)} \beta_{X-k}^{(1)} \sigma_{e_1}^2 t$$
$$= k\beta_X^{(0)} \beta_{X-k}^{(0)} \sigma_{e_0}^2 + (\beta_X^{(1)} \beta_{X-k}^{(1)} \sigma_{e_1}^2 + \beta_X^{(0)} \beta_{X-k}^{(0)} \sigma_{e_0}^2) t. \quad (17)$$

The constraint $T > X + 2$ ensures that $s = t + k + 1, k = 1, \ldots, X$, takes at least two different values that are less than or equal to T. Hence, for $X - k$ fixed (17) holds for at least two different values of t. Therefore, the assumption $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$ implies $\beta_X^{(0)} \beta_{X-k}^{(0)} \sigma_{e_0}^2 = \tilde{\beta}_X^{(0)} \tilde{\beta}_{X-k}^{(0)} \tilde{\sigma}_{e_0}^2$ for $k = 1, \ldots, X$. Together with (16) we can conclude, as in the proof of Theorem 3.1, that either $\text{sgn}(\beta_x^{(0)}) = \text{sgn}(\tilde{\beta}_x^{(0)})$, $\forall x \in \{0, \ldots, X\}$ or $\text{sgn}(\beta_x^{(0)}) = -\text{sgn}(\tilde{\beta}_x^{(0)})$, $\forall x \in \{0, \ldots, X\}$ whenever $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$. We then obtain as in the proof of Theorem 3.1 that $\text{sgn}(\tilde{\beta}_x^{(0)}) = \text{sgn}(\beta_x^{(0)})$ and $c = 1$, which implies $\tilde{\beta}_x^{(0)} = \beta_x^{(0)}$ and $\tilde{\sigma}_{e_0}^2 = \sigma_{e_0}^2$. In view of expression (8) this result together with $g_{\theta}(x, y, s, t) = g_{\tilde{\theta}}(x, y, s, t)$ entails the restriction

$$\tilde{\beta}_x^{(1)} \tilde{\beta}_y^{(1)} \tilde{\sigma}_{e_1}^2 = \beta_x^{(1)} \beta_y^{(1)} \sigma_{e_1}^2, \ \forall x, y = 0, \ldots, X. \quad (18)$$

Analogously to the case of $\tilde{\beta}_x^{(0)}$ and $\tilde{\sigma}_{e_0}^2$, only alternative parameterizations that satisfy $|\tilde{\beta}_x^{(1)}| = c|\beta_x^{(1)}|$, $x = 0, \ldots, X$ are compatible with (18). The same reasoning as in the case of $\tilde{\beta}_x^{(0)}$ and $\tilde{\sigma}_{e_0}^2$ shows $\tilde{\beta}_x^{(1)} = \beta_x^{(1)}$ and $\tilde{\sigma}_{e_1}^2 = \sigma_{e_1}^2$. Hence, $(\tilde{\beta}^{(0)}, \tilde{\beta}^{(1)}, \tilde{\sigma}_{e_0}^2, \tilde{\sigma}_{e_1}^2, \tilde{\sigma}_\epsilon^2) = (\beta^{(0)}, \beta^{(0)}, \sigma_{e_0}^2, \sigma_{e_1}^2, \sigma_\epsilon^2)$.

Consider now the expected values. Since $(\tilde{\beta}^{(0)}, \tilde{\beta}^{(1)}) = (\beta^{(0)}, \beta^{(1)})$ Eq. (7) boils down to

$$\alpha_x - \tilde{\alpha}_x + \beta_x^{(0)}(\mu_0 - \tilde{\mu}_0)(X - x)$$
$$= -(\beta_x^{(0)}(\mu_0 - \tilde{\mu}_0) + \beta_x^{(1)}(\mu_1 - \tilde{\mu}_1))t. \quad (19)$$

Because the left-hand side does not depend on t and $T \geq 4$, both sides of the above equation must equal zero. Hence, $\beta_x^{(0)}(\mu_0 - \tilde{\mu}_0) + \beta_x^{(1)}(\mu_1 - \tilde{\mu}_1) = 0$. If there is a $\bar{x} \in \{0, \ldots, X\}$ such that $\beta_{\bar{x}}^{(0)} = 0$ but $\beta_{\bar{x}}^{(1)} \neq 0$ or such that $\beta_{\bar{x}}^{(1)} = 0$ but $\beta_{\bar{x}}^{(0)} \neq 0$ we obtain $\mu_1 = \tilde{\mu}_1$ or $\mu_0 = \tilde{\mu}_0$, respectively. Indeed if $\beta_{\bar{x}}^{(0)} = 0$ we must have $\mu_1 = \tilde{\mu}_1$, because $\beta_{\bar{x}}^{(1)} \neq 0$. Under the constraints there must be at least one $\beta_x^{(0)} \neq 0$ which then implies $\mu_0 = \tilde{\mu}_0$. Similarly if $\beta_{\bar{x}}^{(1)} = 0$ but $\beta_{\bar{x}}^{(0)} \neq 0$. Hence, it remains to consider that $\forall x \in \{0, \ldots, X\}$ we have either $\beta_x^{(0)} = 0 \land \beta_x^{(1)} = 0$ or $\beta_x^{(0)} \neq 0 \land \beta_x^{(1)} \neq 0$. Whenever $\beta_x^{(0)} \neq 0$ we can rewrite $\beta_x^{(0)}(\mu_0 - \tilde{\mu}_0) + \beta_x^{(1)}(\mu_1 - \tilde{\mu}_1) = 0$ as $\tilde{\mu}_0 = \mu_0 + (\beta_x^{(1)}/\beta_x^{(0)})(\mu_1 - \tilde{\mu}_1)$. Here $\tilde{\mu}_0$ does not depend on x, implying that $\beta_x^{(1)}/\beta_x^{(0)}$ is constant as a function of x, i.e. $\beta_x^{(1)} = d\beta_x^{(0)}$, $\forall x$, for some constant d. However only $d = 1$ satisfies the identifying restrictions made in the model, i.e. $\beta_x^{(1)} = \beta_x^{(0)}$, $\forall x$. But there is no $\theta \in \Theta_{(0,1,0)\times(0,1,0)}^{ct}$ such that $\beta^{(0)} = \beta^{(1)}$, which now implies $\tilde{\mu}_1 = \mu_1$ and $\tilde{\mu}_0 = \mu_0$. Given these latter results, the left-hand side of (19) reduces to $\alpha_x - \tilde{\alpha}_x = 0$ whose only solution is $\alpha_x = \tilde{\alpha}_x$. $\square$

A.2. ARIMA(1,1,0) plug-in Lee–Carter model and proof of Theorem 4.1

Before presenting the proof of Theorem 4.1 we give some background information on the autoregressive integrated moving average process with parameter (1,1,0). If (y_t) is such a process we have with L being the lag-operator $(1 - \rho L)\Delta y_t - \mu = e_t$ which can be written as

$$y_t = y_0 + \mu t + \sum_{s=1}^{t} \sum_{\ell=0}^{\infty} \rho^\ell e_{s-\ell}; \quad (20)$$

see, for instance, Brockwell and Davis (2006, Section 9.1).

For the Lee–Carter model

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \quad x = 1, \ldots, X, \ t = 1, \ldots, T,$$

we obtain upon replacing (κ_t) by an autoregressive integrated moving average process with parameter $(1,1,0)$

$$\log(m_{x,t}) = \alpha_x + \beta_x c + \beta_x \mu t + \beta_x \sum_{s=1}^{t}\sum_{\ell=0}^{\infty} \rho_1^{\ell} e_{s-\ell} + \epsilon_{x,t},$$

where $c = \kappa_0$. This is Eq. (9) in Section 4. The first moments and the covariance structure of this model are directly obtained from the below derivations for (y_t).

The expected value of the process defined by (20) is simply given by $\mathbb{E}(y_t) = y_0 + \mu t$. Let $t \wedge q = \min\{t, q\}$ and $t \vee q = \max\{t, q\}$. Then, for the second moments of this process we find

$$\mathbb{C}\mathrm{ov}(y_t, y_q) = \mathbb{E}\left[\left(\sum_{s=1}^{t}\sum_{\ell=0}^{\infty} \rho^{\ell} e_{s-\ell}\right)\left(\sum_{r=1}^{q}\sum_{k=0}^{\infty} \rho^{k} e_{r-k}\right)\right]$$

$$= \sum_{s=1}^{t}\sum_{r=1}^{q}\sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} \rho^{k+\ell}\mathbb{E}(e_{s-\ell}e_{r-k})$$

$$= \sum_{s=1}^{t \wedge q}\sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} \rho^{k+\ell}\mathbb{E}(e_{s-\ell}e_{s-k})$$

$$+ 2\sum_{s=2}^{t \wedge q}\sum_{r=1}^{s-1}\sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} \rho^{k+\ell}\mathbb{E}(e_{s-\ell}e_{r-k})$$

$$+ \sum_{s=(t \wedge q)+1}^{t \vee q}\sum_{r=1}^{t \wedge q}\sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} \rho^{k+\ell}\mathbb{E}(e_{s-\ell}e_{r-k})$$

$$= I + II + III.$$

Using Hamilton (1994, Eqs. (3.4.4) and (3.4.5)) for the covariances of an autoregressive process with lag 1, we obtain

$$I = (t \wedge q)\frac{\sigma^2}{1 - \rho^2}$$

and

$$II = 2\frac{\sigma^2}{1 - \rho^2}\sum_{s=2}^{t}\sum_{r=1}^{s-1} \rho^{s-r}.$$

Using the geometric series formula twice the sum of these two expressions can be simplified to yield

$$I + II = (t \wedge q)\frac{\sigma^2}{1 - \rho^2} + 2\frac{\sigma^2}{1 - \rho^2}$$
$$\left(\frac{\rho^{t \wedge q+1} - \rho}{(\rho - 1)^2} + \frac{(t \wedge q)\rho(1 - \rho)}{(\rho - 1)^2}\right)$$

$$= \sigma^2\frac{(\rho^2 - 1)(t \wedge q) - 2\rho(\rho^{t \wedge q} - 1)}{(\rho - 1)^3(\rho + 1)}$$

$$= (t \wedge q)\frac{\sigma^2}{(1 - \rho)^2} - 2\sigma^2\rho\frac{(\rho^{t \wedge q} - 1)}{(\rho - 1)^3(\rho + 1)}.$$

With regards to III, we can apply the geometric series formula again to obtain

$$III = \sigma^2\rho\frac{\rho^{t \wedge q} - \rho^{t \vee q} + \rho^{|t-q|} - 1}{(\rho - 1)^3(\rho + 1)}.$$

Plugging this result into the main expression of the covariance and re-arranging, we obtain

$$\mathbb{C}\mathrm{ov}(y_t, y_q) = (t \wedge q)\frac{\sigma^2}{(1 - \rho)^2}$$
$$- \sigma^2\frac{\rho(\rho^{t \vee q} - \rho^{|t-q|} + \rho^{t \wedge q} - 1)}{(\rho - 1)^3(\rho + 1)}. \tag{21}$$

We now turn to the proof of Theorem 4.1.

Proof of Theorem 4.1. We start by looking at the variances of (9). Combining it with I and II from the above we have

$$\mathbb{V}\mathrm{ar}(\log(m_{x,t})) = \beta_x^2\sum_{\ell=1}^{t}\sum_{s=1}^{t}\sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \rho^{k+r}\mathbb{E}(e_{\ell-k}e_{s-r}) + \sigma_{\epsilon}^2$$

$$= \beta_x^2\left(t\frac{\sigma_e^2}{1 - \rho^2} + 2\frac{\sigma_e^2}{1 - \rho^2}\sum_{\ell=2}^{t}\sum_{k=1}^{\ell-1} \rho^{\ell-k}\right) + \sigma_{\epsilon}^2,$$

which for fixed x is the sum of a constant and two terms that depend on t. Consequently, if we have a parameter $\tilde{\boldsymbol{\theta}}$ that leads to the same covariance structure as $\boldsymbol{\theta}$ we can write

$$\beta_x^2\left(t\frac{\sigma_e^2}{1 - \rho^2} + 2\frac{\sigma_e^2}{1 - \rho^2}\sum_{\ell=2}^{t}\sum_{k=1}^{\ell-1} \rho^{\ell-k}\right)$$
$$= \tilde{\beta}_x^2\left(t\frac{\tilde{\sigma}_e^2}{1 - \tilde{\rho}^2} + 2\frac{\tilde{\sigma}_e^2}{1 - \tilde{\rho}^2}\sum_{\ell=2}^{t}\sum_{k=1}^{\ell-1} \tilde{\rho}^{\ell-k}\right) + z \tag{22}$$

or

$$t\left(\beta_x^2\frac{\sigma_e^2}{1 - \rho^2} - \tilde{\beta}_x^2\frac{\tilde{\sigma}_e^2}{1 - \tilde{\rho}^2}\right) = -2\left(\beta_x^2\frac{\sigma_e^2}{1 - \rho^2}\sum_{\ell=2}^{t}\sum_{k=1}^{\ell-1} \rho^{\ell-k}\right.$$
$$\left. - \tilde{\beta}_x^2\frac{\tilde{\sigma}_e^2}{1 - \tilde{\rho}^2}\sum_{\ell=2}^{t}\sum_{k=1}^{\ell-1} \tilde{\rho}^{\ell-k}\right) + z \tag{23}$$

with $z = \tilde{\sigma}_{\epsilon}^2 - \sigma_{\epsilon}^2$. Because the left-hand side of (23) is an affine linear function in t taking the second differences of both sides yields

$$\frac{\beta_x^2\sigma_e^2}{1 - \rho}\rho^{t-1} - \frac{\tilde{\beta}_x^2\tilde{\sigma}_e^2}{1 - \tilde{\rho}}\tilde{\rho}^{t-1} = 0, \; t = 3, \ldots, T \tag{24}$$

which, unless $\beta_x = \tilde{\beta}_x = 0$ or $\rho = \tilde{\rho} = 0$, can only be zero if $\tilde{\rho} = \rho$ and $\tilde{\beta}_x^2\tilde{\sigma}_e^2 = \beta_x^2\sigma_e^2$ as $T \geq 4$. If $\beta_x = \tilde{\beta}_x = 0$ for some x we know that under the constraint $\sum_{x=0}^{X}\beta_x = 1$ there must be at least one $\bar{x} \in \{0, \ldots, X\}$ such that $\beta_{\bar{x}} \neq 0$. From Eq. (24) with $x = \bar{x}$ we then obtain $\rho = \tilde{\rho}$. If $\rho = \tilde{\rho} = 0$ we have the model considered in Theorem 3.1 and we can proceed as we did there to show that $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. To conclude Eq. (24) implies $\tilde{\rho} = \rho$ and $\tilde{\beta}_x^2\tilde{\sigma}_e^2 = \beta_x^2\sigma_e^2$, $x = 0, \ldots, X$, and we can restrict ourselves to the case $\rho = \tilde{\rho} \neq 0$.

As in the proof of Theorem 3.1 identifiability now follows directly for $X = 0$. For $X \geq 1$, we only need to consider $x \in \check{X}$ with $\check{X}$ as defined in the aforementioned proof. Consider now $\mathbb{C}\mathrm{ov}(\log(m_{x,s}), \log(m_{y,T}))$ for $x \neq y$. Combining (9) and (21)

$$\mathbb{C}\mathrm{ov}(\log(m_{x,s}), \log(m_{y,T}))$$
$$= \sigma_e^2\beta_x\beta_y\left(\frac{s}{(1 - \rho)^2} - \frac{\rho(\rho^T + \rho^s - \rho^{T-s} - 1)}{(\rho - 1)^3(\rho + 1)}\right)$$
$$x, y \in \check{X}, \; x \neq y, \; s \in \{1, \ldots, T\}. \tag{25}$$

First, we need to make sure that there is no ρ, $\rho \neq 0$, such that the term in parentheses is zero for all $s = 1, \ldots, T - 1$. If this is the case the first differences of the expression with respect to s are zero as well, i.e.

$$\frac{1}{(1 - \rho)^2} - \frac{\rho(\rho^s(1 - \rho^{-1}) + \rho^{T-s}(\rho - 1))}{(\rho - 1)^3(\rho + 1)} = 0.$$

Solving for the terms that involve s yields

$$\rho^s(1 - \rho^{-1}) + \rho^{T-s}(\rho - 1) = \rho^{-1}((1 - \rho)^{-2})(\rho - 1)^3(\rho + 1),$$

where the exact expression on the right-hand side can be disregarded. Most important is that the left-hand side must be constant

for all $s = 1, \dots, T - 1$. To check this restriction, we simply set equal the cases $s = 1$ and $T - 1$, yielding

$$\rho(1 - \rho^{-1}) + \rho^{T-1}(\rho - 1) = \rho^{T-1}(1 - \rho^{-1}) + \rho(\rho - 1)$$
$$\Leftrightarrow 1 = \rho^{T-2}.$$

Obviously, for $T > 2$ the only solutions to this equality are $\rho = 0$ and $\rho = 1$ as well as $\rho = -1$ for odd T. However, none of these values are elements of the parameter space defined below Eq. (9). Hence, for $T > 2$ there are no permissible values for ρ such that the term in parentheses in Eq. (25) equals zero.

Given the knowledge that the covariances are non-zero (if $x, y \in \check{X}$) and that $\tilde{\rho} = \rho$, a valid alternative parametrization of the second moments must hence satisfy $\sigma_e^2 \beta_x \beta_y = \tilde{\sigma}_e^2 \tilde{\beta}_x \tilde{\beta}_y$. This restriction is identical to that found in the proof of Theorem 3.1 and it can be proceeded in exactly the same way as done there to obtain $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\tilde{\sigma}_e^2 = \sigma_e^2$. We finish the proof by reconsidering (22). Given the results established in this proof, the equality holds only if $z = 0$ which implies $\tilde{\sigma}_\epsilon^2 = \sigma_\epsilon^2$.

It remains to consider the expected values. From (9), we have

$$\mathbb{E}(\log(m_{x,t})) = \alpha_x + \beta_x c + \beta_x \mu t.$$

This expression is identical to Eq. (4), which contained the expected values of the ARIMA(0,1,0) model. Together with $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$, this provides the same starting point for showing that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ and $\tilde{\mu} = \mu$ as done in the proof of Theorem 3.1. $\square$

A.3. ARIMA(0,1,1) plug-in Lee–Carter model and proof of Theorem 4.2

The ARIMA(0,1,1) model is given by $\Delta y_t - \mu = (1 + \phi_1 L)e_t$ so that the process in levels is expressed by

$$y_t = y_0 + \mu t + \sum_{s=1}^{t}(e_s + \phi_1 e_{s-1}).$$

Given that the expected value of this process is $\mathbb{E}(y_t) = y_0 + \mu t$, the covariances are given by

$$\mathbb{C}\text{ov}(y_t, y_q) = \mathbb{E}\left(\sum_{s=1}^{t}\sum_{r=1}^{q}\right)(e_s + \phi_1 e_{s-1})(e_r + \phi_1 e_{r-1}).$$

Again, let $t \vee q = \max\{t, q\}$ and $t \wedge q = \min\{t, q\}$. Furthermore, define $A(s, r) = e_s e_r + \phi_1 e_{s-1} e_r + \phi_1 e_s e_{r-1} + \phi_1^2 e_{s-1} e_{r-1}$. We can decompose the above sum into

$$\mathbb{C}\text{ov}(y_t, y_q) = \sum_{s=1}^{t \wedge q} \mathbb{E}(A(s, s)) + \sum_{s=1}^{t \wedge q}\sum_{r \neq s}^{t \wedge q} \mathbb{E}(A(s, r))$$
$$+ \sum_{s=(t \wedge q)+1}^{t \vee q}\sum_{r=1}^{t \wedge q} \mathbb{E}(A(s, r))$$
$$= I + II + III.$$

For I we directly obtain $I = \sigma_e^2(t \wedge q)(1 + \phi_1^2)$. III reduces to

$$III = \mathbb{1}_{\{t \neq q\}}(t, q)\sigma_e^2 \phi_1$$

because $\mathbb{E}(A((t \wedge q) + 1, t \wedge q))$ is the only nonzero term involved. Lastly, $II = 2((t \wedge q) - 1)\sigma_e^2 \phi_1$ since only $\mathbb{E}(A(s, s - 1))$ and $\mathbb{E}(A(r - 1, r))$ are nonzero. Consequently, $\mathbb{C}\text{ov}(y_t, y_q) = \sigma_e^2(t \wedge q)(1 + \phi_1^2) + 2((t \wedge q) - 1)\sigma_e^2 \phi_1 + \mathbb{1}_{\{t \neq q\}}(t, q)\sigma_e^2 \phi_1$ which can be expressed alternatively as

$$\mathbb{C}\text{ov}(y_t, y_q) = \sigma_e^2\left[(t \wedge q)(\phi_1 + 1)^2 - (1 + \mathbb{1}_{\{t=q\}}(t, q))\phi_1\right].$$

Using this expression within the Lee–Carter model yields

$$\mathbb{C}\text{ov}(\log(m_{x,q}), \log(m_{y,t}))$$
$$= \beta_x \beta_y \sigma_e^2\left[(t \wedge q)(\phi_1 + 1)^2 - (1 + \mathbb{1}_{\{t=q\}}(t, q))\phi_1\right]$$
$$+ \mathbb{1}_{\{t=q\}}(t, q)\mathbb{1}_{\{x=y\}}(x, y)\sigma_\epsilon^2. \tag{26}$$

This result allows us to turn to the proof of Theorem 4.2.

Proof of Theorem 4.2. Consider the covariance structure (26). For $s = t$ and fixed x, y such that $x \neq y$ the above expression reduces to

$$\mathbb{C}\text{ov}(\log(m_{x,t}), \log(m_{y,t})) = \beta_x \beta_y \sigma_e^2\left[t(\phi_1 + 1)^2 - 2\phi_1\right],$$

which is an affine function in t. Since this function can be scaled arbitrarily by a factor z that is compensated by $\tilde{\beta}_x \tilde{\beta}_y \tilde{\sigma}_e^2$, we have the restrictions

$$z(1 + \tilde{\phi}_1)^2 = (1 + \phi_1)^2$$
$$z\tilde{\phi}_1 = \phi_1,$$

which yield

$$\frac{(1 + \phi_1)^2}{(1 + \tilde{\phi}_1)^2} = \frac{\phi_1}{\tilde{\phi}_1}.$$

Solving this equation in $\tilde{\phi}_1$ for a given ϕ_1 we find that either $\tilde{\phi}_1 = \phi_1$ or $\tilde{\phi}_1 = 1/\phi_1$ for $\phi_1 \neq 0$. However, the second solution does not belong to $\Theta_{(0,1,1)}$. Consequently, $\tilde{\phi}_1 = \phi_1$. The remainder of this proof is identical to the proof of Theorem 3.1 as the problem now reduces to $\beta_x \beta_y \sigma_e^2 = \tilde{\beta}_x \tilde{\beta}_y \tilde{\sigma}_e^2$. $\square$

References

Anderson, T.W., Rubin, H., 1956. Statistical inference in factor analysis. in: J. Neyman (ed.), in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V. University of California Press, Berkeley, pp. 114–150.

Booth, H., 2006. Demographic forecasting: 1908 to 2005 in review. Int. J. Forecast. 22, 547–581.

Booth, H., Tickle, L., 2008. Mortality modelling and forecasting: a review of methods. Ann. Actuar. Sci. 3, 3–43.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J., 1997. Actuarial Mathematics, second ed.,Society of Actuaries, Schaumburg.

Brockwell, P.J., Davis, R.A., 2006. Time Series: Theory and Methods, second ed.. Springer Science Business Media, New York.

Brouhns, N., Denuit, M., van Keilegom, I., 2005. Bootstrapping the Poisson log-bilinear model for mortality forecasting. Scand. Actuar. J. 3, 212–224.

Cairns, A.J.G., Blake, D., Dawson, P., Dowd, K., 2005. Pricing the risk on longevity bonds. Life Pensions 41–44 October.

Cairns, A.J.G., Blake, D., Dowd, K., 2008. Modelling and management of mortality risk: a review. Scand. Actuar. J. 2008, 79–113.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. N. Am. Actuar. J. 13, 1–35.

Coelho, E., Nunes, L., 2011. Forecasting mortality in the event of a structural change, J. R. Stat. Soc. Ser. A, 174, 713–734.

Currie, I.D., 2016. On fitting generalized linear and non-linear models of mortality. Scand. Actuar. J. 2016, 356–383.

De Jong, P., Tickle, L., 2006. Extending Lee–Carter mortality forecasting. Math. Popul. Stud. 13, 1–18.

Fung, M.C., Peters, G.W., Shevchenko, P.V., 2016. A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. Ann. Actuar. Sci.. http://dx.doi.org/10.1017/S1748499517000069.

Girosi, F., King, G., (2007). Understanding the Lee–Carter Mortality Forecasting Method. Copy at http://j.mp/lTXlGe.

Haberman, S., Renshaw, A., 2008. Mortality, longevity and experiments with the Lee–Carter model. Lifetime Data Anal. 14, 286–315.

Haberman, S., Renshaw, A., 2009. On age-period-cohort parametric mortality rate projections. Insurance Math. Econom. 45, 255–270.

Haberman, S., Renshaw, A., 2011. A comparative study of parametric mortality projection models. Insurance Math. Econom. 48, 35–55.

Hald, A., 2003. A History of Probability and Statistics and their Applications before 1750. John Wiley & Sons, Inc., Hoboken, New Jersey.

Heaton, C., Solo, V., 2004. Identification of causal factor models of stationary time series. Econom. J. 7, 618–627.

Hunt, A., Villegas, A.M., 2015. Robustness and convergence in the Lee–Carter model with cohort effects. Insurance Math. Econom. 64, 186–202.

Kuang, D., Nielsen, B., Nielsen, J.-P., 2008a. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95, 987–991.

Kuang, D., Nielsen, B., Nielsen, J.-P., 2008b. Identification of the age-period-cohort model and the extended chain-ladder model. Biometrika 95, 979–986.

Kuang, D., Nielsen, B., Nielsen, J.-P., 2011. Forecasting in an extended chain-ladder-type model. J. Risk Insur. 78, 345–359.

Lee, R.D., Carter, L.R., 1992. Modeling and forecasting U. S. mortality. J. Amer. Statist. Assoc. 87, 659–671.

Leng, X., Peng, L., 2016. Inference pitfalls in Lee–Carter model for forecasting mortality. Insurance Math. Econom. 70, 58–65.

Nielsen, B., Nielsen, J.-P., 2014. Identification and forecasting in mortality models. Sci. World J. 2014, Article ID 347043, http://dx.doi.org/10.1155/2014/347043.

Rencher, A.C., 2002. Methods of Multivariate Analysis, second ed.,John Wiley & Sons, Inc., New York.

Renshaw, A.E., Haberman, S., 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance Math. Econom. 38, 556–570.

Yang, B., Li, J., Balasooriya, U., 2016. Cohort extensions of the Poisson common factor model for modelling both genders jointly. Scand. Actuar. J. 2016 (2), 93–112.