Insurance: Mathematics and Economics 75 (2017) 117–125

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

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

ARTICLE INFO

Article history:
Received December 2016
Received in revised form April 2017
Accepted 24 April 2017
Available online 19 May 2017

Keywords:
Time series model
Identifiability
Lee–Carter model
Plug-in Lee–Carter model
Age–period model
Age–period–cohort model

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.

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

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

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.

http://dx.doi.org/10.1016/j.insmatheco.2017.04.006
0167-6687/

defined by

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

Here $\boldsymbol{\iota} = (\iota_{1-X}, \dots, \iota_T)$ 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 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ (model (1)) or $\boldsymbol{\alpha}$, $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$ (model (2)), respectively, are estimated. In the second step a time series model that allows for forecasting is fitted to the estimated $\hat{\boldsymbol{\kappa}}$-vector or to the estimated $\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\iota}}$-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 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ (model (1)) or $\boldsymbol{\alpha}$, $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$ (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 $\boldsymbol{\kappa}$, or $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$, 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 $\boldsymbol{\kappa}$ or identifiable time series models for $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$, 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 $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ 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 $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$. More precisely, we first look at the age–period model (1) if a random walk is used to model the factor scores for $\boldsymbol{\kappa}$. Afterwards we analyse the age–period–cohort Lee–Carter model (2) if two independent random walks are used to model the factor scores for $\boldsymbol{\kappa}$ and $\boldsymbol{\iota}$. 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 $(\kappa_1, \dots, \kappa_{\check{T}})$ 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 $(\kappa_1, \dots, \kappa_{\check{T}})$ which is given to us. This is different from increasing X, because $\boldsymbol{\beta}$ 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

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