Computational Statistics and Data Analysis 83 (2015) 65–81

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Modeling and forecasting duration-dependent mortality rates

Marcus C. Christiansen, Andreas Niemeyer *, Lucia Teigiszerová

Institut für Versicherungswissenschaften, Universität Ulm, D-89069 Ulm, Germany

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 18 August 2014
Accepted 19 September 2014
Available online 28 September 2014

Keywords:
Stochastic mortality
Duration dependence
Disability insurance
Lee–Carter model
Lexis diagram

ABSTRACT

Mortality data of disabled individuals are studied and parametric modeling approaches for the force of mortality are discussed. Empirical observations show that the duration since disablement has a strong effect on mortality rates. In order to incorporate duration effects, different generalizations of the Lee–Carter model are proposed. For each proposed model, uniqueness properties and fitting techniques are developed, and parameters are calibrated to mortality observations of the German Pension Insurance. Difficulties with coarse tabulation of the empirical data are solved by an age–period–duration Lexis diagram. Forecasting is demonstrated for an exemplary model, leading to the conclusion that duration dependence should not be neglected. While the data shows a clear longevity trend with respect to age, significant fluctuations but no systematic trend is observed for the duration effects.
© 2014 Elsevier B.V. All rights reserved.

1. Introduction

For the forecasting of future mortality rates, stochastic models have become more and more popular in demographic research, since they allow for describing not only a future best estimate development but also the uncertainty around a best estimate forecast. Applications range from governmental planning to private business, in particular insurance. In a seminal paper, Lee and Carter (1992) suggested a parametric model that allows stochastic forecasting of a population's mortality with respect to age and calendar year. Apart from being the first model of its kind, the popularity of this approach comes from (a) the reasonable fit for most Western countries, (b) the intuitive interpretation of the models parameters and (c) the simplicity and ease of use in practice. A number of authors further developed the model of Lee and Carter (1992), for example Brouhns et al. (2002), Cairns et al. (2006), Hyndman and Ullah (2007), and Renshaw and Haberman (2006).

In the present paper we study the mortality of disabled people. The results have several applications, e.g., the planning of health care infrastructure and the risk management of disability insurance portfolios. Based on empirical data, we suggest and discuss several stochastic forecasting models. All our models are extensions of the Lee–Carter approach. This way we retain the major advantages of the Lee–Carter approach, in particular the simple calibration, the fact that all parameters have an economic meaning, and the intuitive forecasting into the future. The most important covariates for the mortality rate of healthy people are age and calendar year. For disabled people, several studies have shown that the duration since disablement is a significant covariate (compare e.g. Segerer, 1993). Our data set confirms this observation. While the Lee–Carter model neglects duration dependence, all our extensions feature this covariate.

Our empirical study is based on a data set from the Research Data Center of the German Pension Insurance (Forschungsdatenzentrum der Rentenversicherung, FDZ-RV, 2011), which contains mortality data for disabled policyholders in the German public disability insurance system in the years 1994–2009. All our proposed parametric models are judged on the basis

* Corresponding author. Tel.: +49 731 5031181; fax: +49 731 5031188.
E-mail addresses: marcus.christiansen@uni-ulm.de (M.C. Christiansen), andreas.niemeyer@uni-ulm.de (A. Niemeyer), lucia.teigiszerova@gmail.com (L. Teigiszerová).

http://dx.doi.org/10.1016/j.csda.2014.09.017
0167-9473/© 2014 Elsevier B.V. All rights reserved.

of this data set. Our results show that it is necessary to consider duration dependence in order to achieve a reasonable fit to the empirical observations.

While the literature offers various Lee–Carter extensions, none of them features duration dependence for disabled people. Pitacco (2012) discusses duration dependence in the mortality of disabled people, looking from the perspective of private disability insurance, but he disregards calendar year dependence. Interestingly, Pitacco (2012) presents a model that links the mortality of disabled people with the mortality of healthy people, supposing some functional dependence. Some of our models discussed later on also have such a structure. Aro et al. (2013) model the termination probabilities of a disability insurance using logistic regression and fit their model to data of a Swedish insurance company. Renshaw and Haberman (2000) describe permanent health insurance with a Poisson regression. Both papers take into account duration dependence, but they do not show how their models can be forecast into the future. The classical Lee–Carter approach uses singular value decomposition in order to decompose the age and time matrix to single age and time vectors. Russolillo et al. (2011) extends the Lee–Carter model by a third factor that relates to the country and uses the Tucker3 method to decompose the age, time, and country tensor to single age, time, and country vectors. Unlike the singular value decomposition, the Tucker3 method is rather tricky and adds new problems. Christiansen et al. (2012) use a multivariate Lee–Carter generalization proposed in Hyndman and Ullah (2007) to model mortality, disability and reactivation in a disability insurance, but they also neglect duration dependence. Another stream of research is to directly model the health status of individuals by proportional hazards models with (time-varying) frailty parameters, see e.g. Lam and Wong (2014). In doing so the observed duration effect is interpreted as a partial cure of the disabled individuals.

The paper is structured as follows. In Section 2 we explain our data set, we show how we calculate the empirical mortality rates, and we give some summary statistics. In Section 3 we fit the classical Lee–Carter model to our data. The main part of the paper is Section 4, where different extensions of the Lee–Carter model are proposed and discussed. In Section 5 we explain how to do forecasts, demonstrating the estimation of confidence intervals for one of our proposed models. Section 6 concludes.

2. Description of the data and estimation of the empirical mortality rates

In this section, we first describe the data set and then how the empirical mortality rates can be estimated from that data. Our data set comes from the Research Data Center of the German Pension Insurance (Forschungsdatenzentrum der Rentenversicherung, FDZ-RV, 2011) and includes 1% of the total portfolio of pensioners at the end of the years 1993–2008 (sample size 282,415) and 10% of the discontinuations of pensioners due to mortality in the years 1994–2009 (sample size 85,131). The 1% subset and the 10% subset were randomly chosen from the total data. We only consider reduced earnings capacity pensions, which offer an income protection in case of partial or total disability. The data contains various covariates, but we consider only the age, the calendar time, and the duration since disablement. Notably, we do not consider the gender. The German Pension Insurance paid the pensions monthly in advance, so that the last annuity within a year was paid at the beginning of December. An insured person was classified as being alive if he survived till the end of November. Analogously, if an insured person died in December, he was counted as dead in January. Therefore, we shift the calendar years by one month such that the year starts at the beginning of December and ends at the end of November.

For the duration d since disablement we consider 7 different classes, where the first class 0 means that the person is up to one year disabled, 1 between one and two years, and 6 between six and seven years. For durations of more than seven years the observed data is scarce, so we exclude it. The data gives the duration only on a yearly time grid and, consequently, we can only study discrete durations. We study policyholders with age x between 40 and 59. For younger ages the data basis is small, and people with age 60 or older can alternatively receive old age pension benefits so that adverse selection effects inhibit the data. In total, the covariates and their domains are

$$x \in \{40, \dots, 59\}, \qquad d \in \{0, \dots, 6\}, \quad \text{and} \quad t \in \{1994, \dots, 2009\}.$$

In the following theoretical discussion we use the general notation $x \in \{x_1, \dots, x_k\}$, $d \in \{d_1, \dots, d_l\}$, $t \in \{t_1, \dots, t_m\}$. We now discuss the estimation of the mortality rates m_{xdt}. We generally assume that m_{xdt} is constant in between integer ages, integer durations and integer years, that is,

$$m_{xdt} = m_{\lfloor x \rfloor \lfloor d \rfloor \lfloor t \rfloor}, \quad x_1 \le x < x_k + 1, d_1 \le d < d_l + 1, t_1 \le t < t_m + 1.$$

Let $\widetilde{L}_{xdt}$ be the number of observed individuals at the beginning of year $t \in \{t_1, \dots, t_m\}$ that are age $x \in \{x_1, \dots, x_k\}$ and belong to duration class $d \in \{d_1, \dots, d_l\}$. Let $\widetilde{D}_{xdt}$ be the number of deaths observed in the time interval $[t, t+1)$ of individuals that have age x at death and that got disabled in year $t-d$. Our data set offers the numbers $\widetilde{L}_{xdt}$ and $\widetilde{D}_{xdt}$ for $x \in \{39, \dots, 59\}$, $t \in \{1994, \dots, 2009\}$, and $d \in \{0, \dots, 6\}$. It is tempting to use the fraction $\widetilde{D}_{xdt}$ over $\widetilde{L}_{xdt}$ as an estimator for m_{xdt}, but this would lead to a systematic error. We now explain how to avoid that systematic error by transforming $\widetilde{D}_{xdt}$ and $\widetilde{L}_{xdt}$ to an adequate death count D_{xdt} and exposure to risk L_{xdt}.

Suppose that for each individual we observe the exact date of birth b, the exact time of disablement $i > b$, and the exact time of death $e > i$. Then $\widetilde{D}_{xdt}$ corresponds in a three-dimensional Lexis diagram with dimensions b, i, d to the solid

$$S(\widetilde{D}_{x,d,t}) = \{(b, i, e) : b < i < e, t \le e < t+1, x \le e-b < x+1, d-1 \le t-i < d\}.$$

The number $\widetilde{L}_{xdt}$ can be interpreted as the exposure that corresponds to the solid

$$S(\widetilde{L}_{x,d,t}) = \{(b, i, e) : b < i < e, t \le e < t+1, x \le t-b < x+1, d \le t-i < d+1\}.$$

Download English Version:

https://daneshyari.com/en/article/414939

Download Persian Version:

https://daneshyari.com/article/414939

Daneshyari.com