Insurance: Mathematics and Economics 52 (2013) 275–285

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

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

Modeling and forecasting mortality rates

Daniel Mitchell *, Patrick Brockett, Rafael Mendoza-Arriaga, Kumar Muthuraman

McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received November 2011
Received in revised form
January 2013
Accepted 4 January 2013

Keywords:
Mortality rates
Statistics
Time series
Mortality forecasting

ABSTRACT

We show that by modeling the time series of mortality rate changes rather than mortality rate levels we can better model human mortality. Leveraging on this, we propose a model that expresses log mortality rate changes as an age group dependent linear transformation of a mortality index. The mortality index is modeled as a Normal Inverse Gaussian. We demonstrate, with an exhaustive set of experiments and data sets spanning 11 countries over 100 years, that the proposed model significantly outperforms existing models. We further investigate the ability of multiple principal components, rather than just the first component, to capture differentiating features of different age groups and find that a two component NIG model for log mortality change best fits existing mortality rate data.

1. Introduction

Modeling and forecasting mortality rates has been an active area of research since Graunt (1662) examined mortality in London to create a warning system related to the onset and spread and decline of the bubonic plague. Gaunt's work showed that while individual life length was uncertain, there was a more predictable pattern of longevity and mortality in groups. Halley (1693) showed how to actually construct a non-deficient mortality table from empirical birth–death data and showed how to perform a life annuity calculation based on this table. Such early tables were empirical and calculation was time consuming. Theoretical mortality modeling began with de Moivre (1725) who postulated a uniform distribution of deaths model, and showed simplified annuity calculation methods. Taking a biological approach to mathematical modeling, Gompertz (1825) assumed that the mortality rate represents the body's propensity to succumb to death and that its inverse was the body's ability to withstand death. Assuming that the change in the body's ability to withstand death is proportional to the ability it has to withstand death to begin with led him to a differential equation whose solution is exponential. Solving this and then for the survival function that corresponds to this the solution yields the double exponential survival curve known as the Gompertz curve, which Gavrilov and Gavrilova (2011) shows to fit data to approximately ages 102–105.

The above mortality models are static, however actual mortality is stochastic and evolves over time. Thus, while the mortality models described above fit data at a fixed point in time, the parameters must be re-fit periodically to accommodate changes in mortality patterns. Moreover, the forecasting of future mortality rates is important and not easily accomplished using static mortality models. Future death rates are important to national governments, corporations, and insurance companies. National governments use forecasts of mortality rates to plan social security and health care programs. Couzin-Frankel (2011) estimate that every additional year of life expectancy in the United States costs the U.S. Social Security Administration \$50 billion. Corporations offering defined benefit pension plans must assure proper funding of future liabilities, however these future liabilities depend on the yet to be observed future mortality rates. A 2006 study, by Pension Capital Strategies and Jardine Lloyd Thompson (2006) in the UK, found that recognizing the underestimation of expected lifetimes in FTSE100 index companies would cause the aggregate deficit in pension reserves to more than double from £46 billion to £100 billion. In 2010 improved life expectancy added £5 billion to corporate pension obligations in the U.K. as seen in Reuters (2010). In the U.S. the level of pension contributions needed for adequate reserving will increase pension liabilities by 5%–10%, as seen in Halonen (2007). Similarly, insurance companies must use mortality forecasts for pricing annuity contracts and to decide on required future cash reserves. In order to identify, elucidate and quantify these trends we must have a model that adequately captures the temporal as well as age specific dynamics of mortality rates.

1.1. Lee–Carter model and inter-temporal evolution of mortality rates

One of the first papers to model the separate effects of current age and year was Lee and Carter (1992). These authors propose a

* Corresponding author. Tel.: +1 512 471 3322.
E-mail address: dan.mitchell@utexas.edu (D. Mitchell).

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2013.01.002

log-bilinear model for mortality rates incorporating both age and year effects. Lee and Carter (1992) has been heavily cited and has been recommended for use by two U.S. Social Security Technical Advisory Panels, it is also used by the U.S. Bureau of the Census as a benchmark model, see for example Hollmann et al. (2000). Moreover since this paper, most other models attempting to assess both time and age evolution of mortality have started with the Lee–Carter framework.

Lee and Carter (1992) model mortality rates for different ages over time by extracting an unobserved state variable from the historical data on mortality rates. This state variable is interpreted as a single temporal mortality evolution index applicable for each age group in the entire population. Since each age group is allowed to respond to the temporal mortality index in different ways, the mortality rate of each age group is some linear function of the temporal mortality index. The mortality index itself is modeled as a Brownian motion with a drift so future predictions of mortality rates can be made by extrapolating the index. Specifically, Lee and Carter (1992) model $m(x, t)$, the central mortality rate of age group x at time t, as a bilinear model for $\ln[m(x, t)]$,

$$m(x, t) = e^{a_x + b_x \kappa_t + \epsilon_{x,t}}. \tag{1.1}$$

Here a_x describes the general shape of the mortality curve, κ_t is the temporal mortality index that captures the evolution of rates over time, and b_x describes each age group's response or susceptibility to the temporal mortality index. If $b_x = 0$ or κ_t is constant, then one returns to static mortality table construction. To estimate the parameters of the model the authors use the singular value decomposition of the matrix of age specific log mortality rates through time to find the matrix of rank one that best approximates the actual log mortality rates. This is numerically equivalent to performing principal component analysis on the covariance matrix of log mortality rate levels.

1.2. Problems with the Lee–Carter model and its variants

Since mortality rates have been trending downwards over at least the last 100 years for all age groups, the Lee–Carter estimation process confounds the first principal component with the time trend. The fact that mortality rates are trending downwards means the covariance matrix of mortality rates vastly overestimates dependence. For example, the covariance over 100 years between the log mortality rates of people aged 5–14 and people aged 65–74 is necessarily very high because early in the time series the mortality rate for both age groups was relatively high compared to their respective means, and later in the time series the mortality rate for both age groups was relatively low. This however does not necessarily mean that if we observe a better than average change in mortality for 65–74 year olds we should expect a better than average change for 5–14 year olds. A cure for cancer would certainly have a large impact on older people and a relatively lesser effect on children, due to the variability in the causes of death for different age groups, but when the cure is found both mortality rates may still decline due to their long term trends. As a result of this phenomenon the fit of the Lee–Carter model can be explained by an exogenous variable.

To illustrate this problem, consider, for example, the consumer price index in Argentina, the miles driven in a particular car, the population of the earth and the GDP of China. These four variables have little relation to each other, but a principal component analysis would certainly show much was "explained" by the first component because all these variables are highly correlated with calendar time. As a simple experiment to illustrate this phenomenon we generated 11 *independent* Brownian motions each with a negative drift and sampled 100 points along each path. The drifts and volatilities of each Brownian motion were randomly chosen from uniform distributions. We then "demeaned" the data and performed a singular value decomposition on the data matrix as would be done in estimating the temporal trend κ_t in Lee–Carter. After repeating the experiment 1000 times we find that according to this matrix decomposition the first singular value, on average, accounted for 99.2% of the variability in the data! Reminding ourselves that this implies that a model with one source of randomness explains 99.2% of the variability, it clearly does not align with the fact that the 11 Brownian motions were generated independently. The only thing these Brownian motions have in common is they all have negative drift, however, following the reasoning of Lee and Carter (1992) we might be led to infer that the first singular value is very informative and we can model all the data by simply modeling this first singular value.

Many papers since Lee and Carter (1992) have tried to improve upon their model by adding more principal components, or a cohort effect, or any range of similar statistical quantities, but they all model the level, and dependence between age groups is modeled using a downward trending temporal trend κ_t. Booth et al. (2006) modify the Lee–Carter model by optimally choosing the time period over which to fit the model and adjust the state variable, κ_t to fit the total number of deaths in each year. DeJong and Tickle (2006) reduce the number of parameters in Lee–Carter to model mortality rates as a smoothed state space model. Yang et al. (2010) use multiple principal components to expand the Lee–Carter model. Chen and Cox (2009) introduce jumps into modeling the state variable, found in Lee and Carter (1992), to increase goodness of fit measures and price insurance linked securities. Deng et al. (2012) use a more advanced jump diffusion model to fit the temporal state variable and Li et al. (2011) identify non-linearities in the temporal state variable. A cohort effect, which incorporates the year of birth into the model, is added to the Lee–Carter model in Renshaw and Haberman (2006). In Booth et al. (2006) the authors compare five variants of the Lee–Carter model with data from several countries. Each of these models are interesting, however all suffer from the same design vulnerability as Lee and Carter (1992) described below, in that they all model the level of log mortality rates and hence misrepresent the temporal dependence structure of mortality rates by age group.

1.3. Outline and contribution

In this paper we build upon the idea of bilinear modeling of age and time from Lee and Carter (1992), however we propose a model that looks at mortality data from a different perspective. We show that by first performing a simple transformation of the data prior to modeling, the subsequent modeling vastly improves our ability to replicate the dynamics of mortality rates through time. This improvement applies not only to the original Lee–Carter model, but also can be used to improve each of the extensions and variants described previously. For forecasting mortality rates we also propose a Normal Inverse Gaussian based mortality index model that is extremely easy to calibrate, has relatively few parameters and performs extremely well. We document this by comparing our model to several other models using several metrics found in literature.

This new model we propose avoids the common problem of modeling log mortality levels. Our model is similar to the Lee–Carter model, however we model changes in log mortality rates rather than levels of log mortality rates. By considering the changes we are able to more accurately capture the dependence structure between ages of mortality and use this to construct a more encompassing model. Referring back to the independent Brownian motions experiment described above, we performed a singular value decomposition on the matrix of differences through