Mathematical Statistics
Stockholm University
Research Report **2019:11**,
http://www.math.su.se

Mortality forecasting using a Lexis based state space model

Patrik Andersson*, Mathias Lindholm†

July 2019

Abstract

A model of mortality is introduced which is based on a data generating process defined in terms of the continuous-time dynamics of a Lexis diagram. Through counting process arguments, the likelihood of death count data sampled at yearly intervals is shown to be equivalent to that of a certain Poisson likelihood. Therefore a hidden Markov model with Poisson distributed observations and a Gaussian state process is introduced. More specifically, the latent mortality rate process is driven by a low-dimensional autoregressive process, where the dimension reduction in a Poisson setting is based on so-called generalized principal component analysis (GPCA). The full likelihood of the Poisson state space model is not analytically tractable, but it is possible to derive explicit sufficient statistics when conditioning on the state of the latent mortality rate process. This makes it possible to estimate the model parameters using the stochastic approximation expectation-maximization (SAEM) algorithm, where sampling is made using particle filter techniques. This circumvents the two-step estimation procedure used for e.g. the Lee-Carter model.

Further, the constructive nature of the introduced model makes it easy to decompose the observed variation in terms of population ("Poisson") variation and variation due to the latent mortality rate process. Since all model parameters are estimated using maximum likelihood theory we argue that it is natural to assess the model performance using logarithmic scores. In particular, we introduce a proper scoring rule based on a transformation of a certain logarithmic score which is closely connected to the maximization step in the SAEM-algorithm. This scoring rule may be seen as a coefficient of determination like measure which can be used for assessing specific age and calendar year model performance, both in-sample and out-of-sample.

The versatility of the model is illustrated on Swedish and US data, where both in-sample and out-of-sample forecast performance is analyzed. We illustrate the convergence of the numerical routines being used and discuss initiation procedures. Further, the numerical illustrations indicate that by not explicitly taking the population part of the variation into account may lead to that too much variation is attributed to the mortality rates, consequently being a potential problem for Lee-Carter type models.

Keywords: Non-linear non-Gaussian state-space-models; Generalized Principal Component Analysis; log-concave likelihood; Stochastic Approximation EM; Particle filter; Mortality forecasting; Hidden Markov model

1 Introduction

Understanding and forecasting mortality is an important part of demographic research and policy making, due to its connection to e.g. pensions, taxation and public health. A closely related area

*Department of Statistics, Uppsala University. Email: patrik.andersson@statistics.uu.se
†Department of Mathematics, Stockholm University. Email: lindholm@math.su.se