Continuous categories in The Lee Carter Model: An application of Gaussian process priors

Brynjólfur Gauti Jónsson

Faculty of Physical Sciences
University of Iceland
2020

CONTINUOUS CATEGORIES IN THE LEE CARTER MODEL: AN APPLICATION OF GAUSSIAN PROCESS PRIORS

Brynjólfur Gauti Jónsson

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Bayesian statistics

Advisor
Professor Birgir Hrafnkelsson

Faculty Representative
XXNN3

M.Sc. committee
XXNN4
XXNN5

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, June 2020

Continuous categories in The Lee Carter Model: An application of Gaussian process priors
Lee Carter
60 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Bayesian statistics

Copyright © 2020 Brynjólfur Gauti Jónsson
All rights reserved

Faculty of Physical Sciences
School of Engineering and Natural Sciences
University of Iceland
Tæknigarður - Dunhaga 5
107, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Brynjólfur Gauti Jónsson, 2020, Continuous categories in The Lee Carter Model: An application of Gaussian process priors, M.Sc. thesis, Faculty of Physical Sciences, University of Iceland.

ISBN XX

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavik, Iceland, June 2020

Dedication

Abstract

bla

Útdráttur

bla

Preface

Contents

List of Figures

List of Tables

Abbreviations

Acknowledgments

Chapter 1

Introduction

1.1 Motivation and Goals

Over the years, progress in all aspects of medicine and public health has led to a steadily declining mortality and longer life expectancy around the world. While these developments are to be celebrated, they entail problems which must be handled. In the public sector, longer life expectancy can lead to higher health care costs as people live longer and the number of elderly patients increases. As for the private sector, insurance funds are heavily affected by changes in mortality, as they base pension annuities on projected mortality rates. If their projections are too high their estimated cost of annuities will be too low, leading to overpayment to clients, and if they are too low they might pay too little annuities, leading to underpayment fewer clients. Thus, it seems there is a need for accurate and precise forecasting of mortality rates.

This project will introduce the Lee-Carter model (Lee and Carter, 1992) for mortality forecasting and a bayesian Markov-Chain Monte-Carlo algorithm that obtains samples from the posterior distribution of all relevant parameters by treating the model as a linear Gaussan state-space model, also known as a dynamic linear model *(DLM)* (West and Harrison, 2006). Two DLMs will be fit, one imitating the classic Lee-Carter model, and one allowing for different observational variances for each age-group.

1.2 Literature Review

1.2.1 Age-Period-Cohort Models

1.2.2 The Lee Carter Model

1.2.3 Stan

1.3 Structure of the Thesis

Chapter 2

Theory and Methods

2.1 Bayesian Inference

This is theory and methods.

2.2 Hierarchical Models

2.3 Gaussian Processes

2.4 Hamiltonian Monte Carlo

2.5 Leave-One-Out Cross-Validation

2.6 Pareto-Smoothed Importance Sampling LOO-CV

Chapter 3

Data and Preliminary Analysis

3.1 Data

The data, coming from Statistics Iceland, contain Icelandic population estimates and number of deaths for all ages $x = [0, 1, \dots, 104]$ and years $t = [1981, 1982, \dots, 2017]$. The ages were reduced into ten groups ($[0, 10)$, ..., $[80, 90)$ and $[90, 104]$) for ease of computation, leaving $10 \times 27 = 270$ age-year-observations of population size and deaths. Figure 1 shows the evolution of population size in the ten age-groups. We can see that the population composition has shifted noticably towards an older population.

Figure 2 shows the evolution of number of deaths within each age-group over time. There has been a noticable decrease in deaths among the youngest part of the population, as well as an increase among the oldest part.

This figure alone is not enough evidence for changes in mortality rates, but when population and deaths are used to calculate observed mortality rates, as is done in figure 3, there is a noticable rate of change in mortality rates, with each age-group seeming to have its own linear trend with added observational noise.

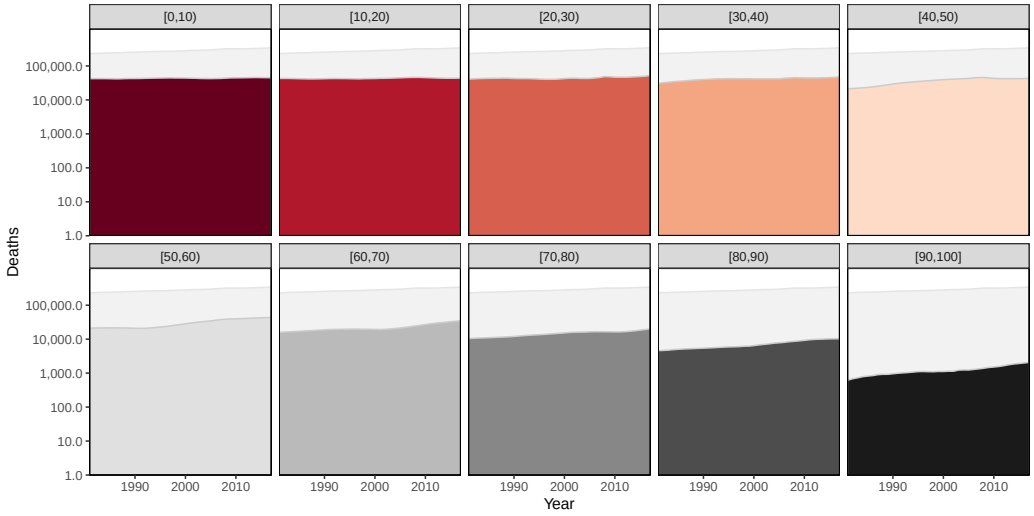

Figure 3.1: Population sizes of year-age-groups, shown in log scale and overlaid on yearly totals for comparison.

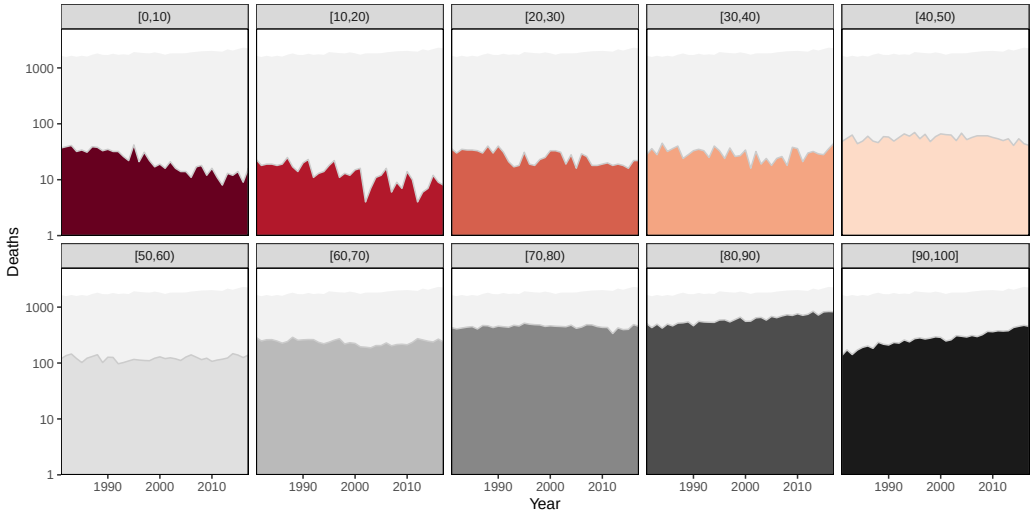

Figure 3.2: Number of deaths by year and age, shown in log scale and overlaid on yearly totals for comparison.

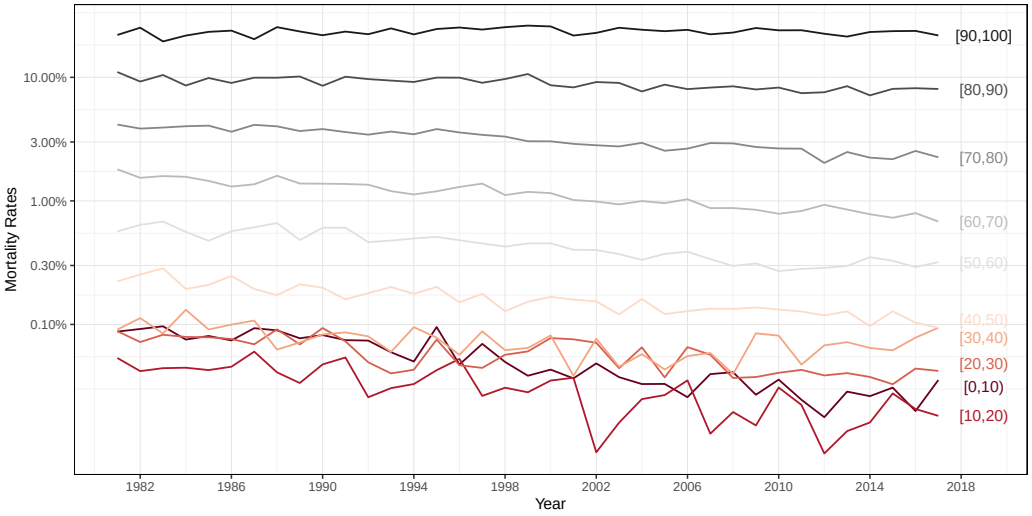

Figure 3.3: Time evolution of mortality rates by age-group, shown on log scale.

Chapter 4

Models and Inference

4.1 The Lee-Carter Model

4.1.1 Parametric Form

There exist many named models for the forecasting of mortality rates, but the most famous one is the Lee-Carter model (Lee and Carter, 1992), named after its authors Ronald D. Lee and Lawrence R. Carter. Let $m(x,t)$ be the mortality rate for age x during year t. The Lee-Carter model can be written

$$\begin{aligned} \ln m(x,t) &= \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \\ \kappa_t &= f(\kappa_{t-1}) + \omega_t, \\ \varepsilon_{x,t} &\sim \mathcal{N}(0, \sigma_\varepsilon^2), \\ \omega_t &\sim \mathcal{N}(0, \sigma_\omega^2) \end{aligned}$$

In this parametrization α_x can be thought of as the mean log-mortality for age-group x. As is a vector and κ_t is a scalar, κ is interpreted as a national mortality index and β_x as the rate-of-change intensity for age-group x. Thus, change in mortality for all age-groups follows the same general pattern, but with different intensities determined by β. Looking back at figure 3 this model fits one intercept and one coefficient of linear change for each age-group, corresponding to the mortality rate at year 1981 and the linear trend, which seems like a logical approximation when inspecting the figure.

In their original paper, Lee and Carter modeled f, the time-trend of κ, as a random walk with drift, i.e.

$$f(\kappa_{t-1}) = \kappa_{t-1} + \theta, \text{ such that}$$
$$\kappa_t = \kappa_{t-1} + \theta + \omega_t$$

but others have included differencing and autoregressive behaviours such that f becomes a second-level ARIMA process

$$(\kappa_t - \mu_\kappa - \beta_\kappa t) = \rho(\kappa_{t-1} - \mu_\kappa - \beta_\kappa(t-1)), \quad 0 < |\rho| < 1,$$

where μ_κ is the expected value of κ_0 and β_κ is the linear trend of κ. Thus the second level ARIMA representation assumes that the observed deviation of κ from its expected value at time t, according to a linear trend, is correlated with that same deviation at time $t-1$ with correlation ρ.

In this project the time-trend of κ is assumed to follow a random walk with drift.

4.1.2 Classical Estimation

The Lee-Carter model as represented thus far is overparametrized, as any change in the estimates of κ can be reversed by adequate changes in α and β. To combat this, Lee and Carter imposed the following constraints to ensure identifiability

$$\sum_i \beta_i = 1, \quad \sum_t \kappa_t = 0.$$

Estimates of the parameters from the Lee-Carter model are commonly obtained by singular value decomposition *(SVD)*. The vector α is set to the average log-mortality for each agegroup and SVD is then applied to the mean-corrected mortality rates $Y - \hat{\alpha} = UDV^T$, where D is a diagonal matrix containing singular values and U and V are orthogonal matrices. β is chosen as the first column of U, and κ is set to be equal to the product of the first singular value, D_{11}, and the first column of V. The correct estimates are then obtained by imposing the necessary parameter restrictions. This is equivalent to performing principal component analysis on a matrix containing the mean-corrected log-mortality rates for all groups at all times and choosing the first principal component.

4.2 A Bayesian State-Space Interpretation

Claudia Pedroza published an article in 2006 where she reformulates the method as a state-space model (Pedroza, 2006):

$$\begin{aligned} y_t &= \ + \ \kappa_t + \varepsilon_t, \qquad \varepsilon \overset{\text{iid}}{\sim} \text{Normal}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}) \\ \kappa_t &= \kappa_{t-1} + \theta + \omega_t, \quad \omega_t \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma_\omega^2) \end{aligned}$$

where y_t is a vector containing the log-mortality rates for all age-groups during year t. As such, the Kalman Filter can be used to estimate and forecast mortality rates. State-space models also have the convenient property of easily handling missing data. In their 2009 book, Dynamic Linear Models with R, Giovanni Petris, Sonia Petorne and Patrizia Campagnoli, refer to the forward-filtering-backwards-sampling method for estimating state-space models using Markov Chain Monte Carlo (Campagnoli et al., 2009). The details of this algorithm are further expanded in the 2006 book Bayesian Forecasting and Dynamic Models by Mike West and Jeff Harrison (West and Harrison, 2006).

4.2.1 Estimation via Gibbs Sampling

Pedroza utilized a Gibbs sampler to obtain samples from the posterior distribution of all parameters. The idea is to alternate between **(1)** drawing the states κ conditioned on the parameters α, β, σ_ε^2, θ, σ_ω^2 and **(2)** drawing the parameters given the states. Let T be the number of timepoints and N be the number of age-groups. Then Y is an $N \times T$ matrix containing all log-mortality rates with y_{xt} equal to the log-mortality rate for age-group x at time t, . First, suitable initial values are chosen for the parameters α, β, κ, θ, σ_ε^2 and σ_ω^2. The sampler is composed of five steps that are then repeated until convergence.

(i) Perform forward filtering and backward sampling to obtain samples from

$$p(\kappa | Y, \alpha, \beta, \sigma_\varepsilon^2, \theta, \sigma_\omega^2).$$

Write $\kappa_t \sim \mathcal{N}(a_t, Q_t)$, and choose initial parameters a_0, R_0. Run the Kalman filter with updating equations

$$\begin{gathered} v_t = y_t - \alpha - \beta a_t, \quad Q_t = \beta R_t \beta^T + \sigma_\varepsilon^2 I_N, \quad K_t = R_t \beta^T Q_t^{-1}, \\ a_{t+1} = a_t + \theta + K_t v_t, \quad R_{t+1} = R_t(1 - K_t \beta) + \sigma_\omega^2, \end{gathered}$$

for $t = 1, \ldots, T$. Next, sample κ as

$$\begin{gathered} \kappa_T | Y, \alpha, \beta, \sigma_\varepsilon^2, \theta, \sigma_\omega^2 \sim \mathcal{N}(a_T, Q_T), \\ \kappa_{t-1} | \kappa_t \sim \mathcal{N}(h_t, H_t), \\ h_t = a_t + B_t(\kappa_{t+1} - a_{t+1}), \quad H_t = Q_t - B_t R_{t+1} B_t^T, \quad B_t = Q_t R_{t+1}^{-1} \end{gathered}$$

(ii) Draw σ_ε^2 from

$$\sigma_\varepsilon^2|Y,\alpha,\beta,\kappa \sim \text{Inv-Gamma}\left(\frac{NT}{2}, \frac{\sum_x\sum_t(y_{xt}-\alpha_x-\beta_x\kappa_t)^2}{2}\right)$$

(iii) Letting y_x be the log-mortality at all time points for age-group x, draw α and β by performing separate linear regressions of y_x on κ for each age group. That is, if X is a $T \times 2$ matrix with first column equal to $[1,\dots,1]$ and the second column equal to κ we sample α_x and β_x from

$$(\alpha_x,\beta_x)|Y,\kappa,\sigma_\varepsilon^2 \sim \mathcal{N}((X^TX)^{-1}X^Ty_x, \sigma_\varepsilon^2(X^TX)^{-1})$$

(iv) Sample θ from

$$\theta|\kappa,\sigma_\omega^2 \sim \mathcal{N}\left(\frac{\kappa_T-\kappa_0}{T}, \frac{\sigma_\omega^2}{T}\right)$$

(v) Draw σ_ω^2 from

$$\sigma_\omega^2|\kappa,\theta \sim \text{Inv-Gamma}\left(\frac{T-1}{2}, \frac{\sum_t(\kappa_t-\kappa_{t-1}-\theta)^2}{2}\right)$$

4.2.2 Extensions

Treating the Lee-Carter model as a Bayesian state-space model makes arbitrary extensions easy. One other model will be fit to the data, allowing for different observational variances in each age-group. The Gibbs sampling algorithm is identical except for simple modifications at steps **(ii)** and **(iii)**:

(ii) $\sigma_{\varepsilon,x}^2$ is calculated separately for each age-group, x.

$$\sigma_{\varepsilon,x}^2|Y,\alpha,\beta,\kappa \sim \text{Inv-Gamma}\left(\frac{T}{2}, \frac{\sum_t(y_{xt}-\alpha_x-\beta_x\kappa_t)^2}{2}\right)$$

(iii) Regression coefficients are now sampled conditional on each age-group's unique observational variance, $\sigma_{\varepsilon,x}^2$.

$$(\alpha_x,\beta_x)|Y,\kappa,\sigma_{\varepsilon,x}^2 \sim \mathcal{N}((X^TX)^{-1}X^Ty_x, \sigma_{\varepsilon,x}^2(X^TX)^{-1})$$

4.2.3 Prior distributions

- α, β
  - α and β are given a flat prior $p(\alpha, \beta) \propto 1$
- σ_ε, σ_ω and $\sigma^2_{\varepsilon,x}$
  - σ_ε, σ_ω and $\sigma^2_{\varepsilon,x}$ are all given a flat prior $p(\sigma) \propto \frac{1}{\sigma}$.
- κ_0, θ
  - κ_0 and θ are given flat priors $p(\kappa_0, \theta) \propto 1$.
  - On the other hand initial values of m_0 and C_0 where $\kappa_0 \sim \mathcal{N}(m_0, C_0)$ are needed, but these are overwritten after the first iteration of the Gibbs sampler.
  - The sampler was tested for several different initial value pairs and found to be insensitive to the choice. The initial values used were $m_0 = 3.74$ and $C_0 = 1$.

4.3 Gaussian Processes

Chapter 5

Results

5.1 Convergence Assesment

This is Results

5.2 Parameters

5.3 Hyperparameters

5.4 Model Checking

5.5 Goodness of Fit

Chapter 6

Conclusions

This is Conclusions

Bibliography

Campagnoli, P., Petris, G., and Petrone, S. (2009). *Dynamic Linear Models with R*. Springer.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Pedroza, C. (2006). A bayesian forecasting model: predicting us male mortality. *Biostatistics*, 7(4):530–550.

West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.