Available online at www.sciencedirect.com
SCIENCE DIRECT®
Insurance: Mathematics and Economics 36 (2005) 260–284
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/ime

Bayesian Poisson log-bilinear mortality projections

Claudia Czado[a], Antoine Delwarde[b], Michel Denuit[b,c,*]

[a] *SCA Zentrum Mathematik, Technische Universität München, D-85748 Garching bei Munich, Germany*
[b] *Institut des Sciences Actuarielles, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*
[c] *Institut de Statistique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

Received September 2004 ; received in revised form January 2005; accepted 21 January 2005

Abstract

Mortality projections are major concerns for public policy, social security and private insurance. This paper implements a Bayesian log-bilinear Poisson regression model to forecast mortality. Computations are carried out using Markov Chain Monte Carlo methods in which the degree of smoothing is learnt from the data. Comparisons are made with the approach proposed by Brouhns et al. [Insur.: Math. Econ. 31 (2002) 373–393; Bull. Swiss Assoc. Actuaries (2002) 105–130], as well as with the original model of Lee and Carter [J. Am. Stat. Assoc. 87 (1992) 659–671].
© 2005 Elsevier B.V. All rights reserved.

Keywords: Projected lifetables; Expected remaining lifetimes; Poisson regression; MCMC

1. Introduction

1.1. Lee–Carter model for mortality projections

Mortality forecasts are used in a wide variety of fields: for health policy making, for directing pharmaceutical research, social security, for retirement fund planning and for life insurance, to name just a few. During the 20th century, it is now well documented that the human mortality globally declined: in most industrialized countries, mortality at adult and old ages reveals decreasing annual death probabilities.

In this paper, we analyze the changes in mortality as a function of both age x and calendar time t. Henceforth, $\mu_x(t)$ will denote the force of mortality at age x and time t. Throughout this paper, we assume that given any integer age x and calendar year t,

* Corresponding author.
E-mail address: denuit@stat.ucl.ac.be (M. Denuit).

0167-6687/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2005.01.001

$$\mu_{x+\xi}(t+\tau) = \mu_x(t) \quad \text{for } 0 \le \xi,\ \tau < 1. \tag{1.1}$$

This is best illustrated with the aid of a coordinate system that has calendar time as abscissa and age as coordinate. Such a representation is called a Lexis diagram after the German demographer who introduced it. Both time scales are divided into yearly bands, which partition the Lexis plane into square segments. Model (1.1) assumes that the mortality rate is constant within each square, but allows it to vary between squares. We denote as D_{xt} the number of deaths recorded at age x during year t, from an exposure-to-risk E_{xt} (that is, E_{xt} is the number of person-years from which D_{xt} occurred).

A powerful and elegant approach to mortality forecasts has been pioneered by Lee and Carter (1992). Those authors proposed a remarkably simple model for mortality projections, specifying a log-bilinear form for the force of mortality $\mu_x(t)$. The method is in essence a relational model

$$\ln \widehat{\mu_x}(t) = \alpha_x + \beta_x \kappa_t + \epsilon_x(t), \tag{1.2}$$

where $\widehat{\mu_x}(t) = \frac{D_{xt}}{E_{xt}}$ denotes the observed force of mortality at age x during year t, the $\epsilon_x(t)$'s are homoskedastic centered error terms and where the parameters are subject to the constraints

$$\sum_t \kappa_t = 0 \quad \text{and} \quad \sum_x \beta_x = 1 \tag{1.3}$$

ensuring model identification.

An important aspect of Lee–Carter methodology is that the time factor κ_t is intrinsically viewed as a stochastic process. Box–Jenkins techniques are then used to estimate and forecast κ_t within an ARIMA time series model. From this forecast of the general level of mortality, the actual age-specific rates are derived using the estimated age effects. This in turn yields projected life expectancies.

For a review of recent applications of the Lee–Carter methodology, we refer the interested readers to Lee (2000). It is worth to mention that the Lee–Carter model is used by the US Census Bureau as a benchmark for their population forecasts, and its use has been recommended by the US Social Security Technical Advisory Panels. It appears to be the determinant method in the literature.

1.2. Poisson log-bilinear model for mortality projections

According to Brillinger (1986) and Alho (2000), the Poisson approximation for the number of deaths occurring in a square of the Lexis diagram is plausible. This led Sithole et al. (2000) and Renshaw and Haberman (2003a,b) to implement an alternative approach to mortality forecasting: calendar time enters the model as a known covariate and a regression model based on heteroskedastic Poisson error structures is used.

A closely related model has been proposed by Brouhns et al. (2002a,b), keeping the Lee–Carter log-bilinear form for the forces of mortality. Specifically, Brouhns et al. (2002a,b) considered that

$$D_{xt} \sim \text{Poisson}(E_{xt}\mu_x(t)) \text{ with } \mu_x(t) = \exp(\alpha_x + \beta_x \kappa_t), \tag{1.4}$$

where the parameters are still subjected to the constraints (1.3).

There is thus a key difference between Renshaw and Haberman (2003a) and Brouhns et al. (2002b) that centres on the interpretation of time: in Brouhns et al. (2002b) time is modeled as a factor and under the approach proposed by Renshaw and Haberman (2003a) is modelled as a known covariate. We believe that the former approach is preferable since we do not constrain ex ante the effect of calendar time to some known functional form. Note that Renshaw and Haberman (2003b) considered both possibilities (factor and covariate) and decided that the one is preferable to the other on the basis of quality of fit.

Instead of resorting to SVD for estimating α_x, β_x and κ_t, Brouhns et al. (2002a,b) estimated the parameters by maximizing the log-likelihood based on model (1.4). As in the Lee–Carter approach, ARIMA models are then used to forecast the κ_t's.

1.3. Scope of the paper

In all the papers mentioned above, the modelling still proceeds in two steps: first the mortality index κ_t is estimated and then it is extrapolated using Box–Jenkins methodology. Possible incoherence may arise from this two-step procedure. In order to avoid this deficiency, we purpose to integrate both steps into a Bayesian model. Bayesian formulations assume some sort of smoothness of age and period effects in order to improve estimation and facilitate prediction. Intuitively, we expect smooth variations of the mortality rates over the Lexis plane. In order to implement this idea, we resort to a Bayesian model in which the prior portion imposes smoothness by relating the underlying mortality rates to each other over the Lexis plane. As a consequence, the rate estimate in each age-year square "borrows strength" from information in adjacent squares. An important advantage of incorporating the idea of smoothness is that it is possible to use the model for purposes of forecasting future mortality rates.

The Bayesian modelling treats all unknown parameters α_x, β_x and κ_t as random variables and derives their distribution conditional upon the known information (E_{xt}, D_{xt}). Until recently, fully Bayesian analyses had been computationally infeasible and approximation methods were often utilized instead. This changed in the early 1990s with computer-intensive Markov Chain Monte Carlo (MCMC) simulation methods (see Chib (2001) for a summary and Gilks et al. (1996) for applications). The Monte Carlo approach allows for inference based on sampling the posterior distribution of the parameters. A particularly attractive feature of this approach is the ease with which we can then explore the uncertainty associated with the estimates and the forecasts.

A Bayesian analysis of mortality has been suggested by several authors. Let us mention the works by Klugman (1989), Yashin et al. (2000) and Dellaportas et al. (2001), to name just a few. A Bayesian treatment of mortality projections has been proposed by Girosi and King (2003). The approach followed by Girosi and King (2003) is nevertheless entirely different from the one adopted in this paper. We refer the reader to the interesting monograph written by Girosi and King (2003) for more details.

1.4. Agenda

Section 2 describes the model and details the prior assumption on each set of parameters. Section 3 derives the MCMC algorithm yielding the a posteriori distribution of the parameters. A numerical illustration is discussed in Section 4, where the results obtained with the methodology developed in this paper are compared with former ones.

By convention, vectors and matrices are denoted by bold lower and upper cases, respectively. Parameters and hyperparameters are denoted by Greek letters. All the vectors are assumed to be column vectors and the superscript $'$ indicates transposition. We denote as $x_{\min}, x_{\min}+1, \ldots, x_{\max}$ the observed age range and as $t_{\min}, t_{\min}+1, \ldots, t_{\max}$ the observed calendar time range. Moreover, $M = x_{\max} - x_{\min} + 1$ is the number of different ages considered in the model, and $T = t_{\max} - t_{\min} + 1$ is the number of calendar years. We denote as $\boldsymbol{I}_M$ (resp. $\boldsymbol{I}_T$) the M-dimensional (resp. T-dimensional) identity matrix. Further, $X \sim \mathcal{N}ormal(m, \sigma^2)$ indicates that the random variable X is normally distributed with mean m and variance σ^2, while $\boldsymbol{X} \sim \mathcal{N}ormal_d(\boldsymbol{m}, \boldsymbol{\Sigma})$ indicates that the d-dimensional random vector $\boldsymbol{X}$ is normally distributed with mean vector $\boldsymbol{m}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ (the subscript d thus indicates the dimension of $\boldsymbol{X}$).

2. Model and prior distributions

2.1. Likelihood function

Let us consider the Poisson log-bilinear model (1.4) supplemented with the constraints (1.3) in order to ensure the identifiability of the model. This model comprises three sets of parameters: $\boldsymbol{\alpha} = (\alpha_{x_{\min}}, \ldots, \alpha_{x_{\max}})'$, $\boldsymbol{\beta} = (\beta_{x_{\min}}, \ldots, \beta_{x_{\max}})'$ and $\boldsymbol{\kappa} = (\kappa_{t_{\min}}, \ldots, \kappa_{t_{\max}})'$. The likelihood function associated with the data points (E_{xt}, D_{xt}), $x = x_{\min}, x_{\min}+1, \ldots, x_{\max}$ and $t = t_{\min}, t_{\min}+1, \ldots, t_{\max}$, is

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) = \prod_x \prod_t \frac{\exp(-E_{xt}\ \exp(\alpha_x + \beta_x\kappa_t))(E_{xt}\ \exp(\alpha_x + \beta_x\kappa_t))^{D_{xt}}}{D_{xt}!}$$
$$\propto \prod_x \prod_t \exp(-E_{xt}\ \exp(\alpha_x + \beta_x\kappa_t) + D_{xt}(\alpha_x + \beta_x\kappa_t)). \tag{2.1}$$

As usual, the first stage of a Bayesian analysis is to specify a prior probability density for the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ involved in the Poisson log-bilinear model. This prior should support the local regularities that are believed to exist.

2.2. Prior distribution for the time index $\boldsymbol{\kappa}$

The time index κ_t represents the time trend. The actual forces of mortality change according to an overall mortality index κ_t modulated by an age response β_x. In the Lee–Carter approach, as well as in its Poisson counterpart, the κ_t's are projected using an ARIMA model. In that respect, a random walk with drift was found the most appropriate for the data analyzed by Lee and Carter (1992). In practice, that simple model for κ_t is used almost exclusively and accounts for nearly all published applications.

Here, we slightly depart from the Lee–Carter approach (which invariably involves an integrated time series). A random walk with drift gives a linear trend for the κ_t's, perturbated by a Gaussian noise. We keep the same idea and we use an autoregressive prior distribution with a linear mean for $\boldsymbol{\kappa}$. More specifically, we consider the model

$$\kappa_t - \gamma_1 - \gamma_2 t = \rho(\kappa_{t-1} - \gamma_1 - \gamma_2(t-1)) + \epsilon_t \text{ for } t = t_{\min}, \ldots, t_{\max} \tag{2.2}$$

with $\kappa_{t_{\min}-1} = \gamma_1 + \gamma_2(t_{\min} - 1)$. The errors $\epsilon_t \sim \mathcal{N}ormal(0, \sigma_\kappa^2)$ are mutually independent. Denoting as $\eta_t = \gamma_1 + \gamma_2 t$ the linear trend for the κ_t's, (2.2) can be re-written as

$$(\boldsymbol{\kappa} - \boldsymbol{X\gamma}) = \boldsymbol{P}(\boldsymbol{\kappa} - \boldsymbol{X\gamma}) + \boldsymbol{\epsilon}$$

with

$$\boldsymbol{P} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \rho & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho & 0 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & t_{\min} \\ \vdots & \vdots \\ 1 & t_{\max} \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\epsilon} \sim \mathcal{N}ormal_T(\boldsymbol{0}, \sigma_\kappa^2 \boldsymbol{I}_T).$$

This leads to the prior $\boldsymbol{\kappa} \sim \mathcal{N}ormal_T(\boldsymbol{X\gamma}, \sigma_\kappa^2 \boldsymbol{Q}^{-1})$ for the time index, with

$$\boldsymbol{Q} = \begin{pmatrix} 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & \vdots \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}.$$

We consider the unknown $\boldsymbol{\gamma}$, ρ and σ_κ^2 involved in the prior distribution of $\boldsymbol{\kappa}$ as hyperparameters. They are treated as random variables with their own prior distributions. The corresponding priors are

$$\boldsymbol{\gamma} \sim \mathcal{N}ormal_2(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0), \qquad \rho \sim \mathcal{N}ormal(0, \sigma_\rho^2) \text{ truncated to the interval } (-1, 1)$$

and we follow the standard practice and choose an inverse gamma prior for σ_κ^2, that is

$$\sigma_\kappa^{-2} \sim \mathcal{G}amma(a_\kappa, b_\kappa).$$

With this choice, $\mathbb{E}[\sigma_\kappa^{-2}] = \frac{a_\kappa}{b_\kappa}$ and $\mathbb{V}\text{ar}[\sigma_\kappa^{-2}] = \frac{a_\kappa}{b_\kappa^2}$. The prior distribution for the vector $\boldsymbol{\kappa}$ is thus determined by the three constants a_κ, b_κ and σ_ρ^2, by the vector $\boldsymbol{\gamma}_0$ and by the matrix $\boldsymbol{\Sigma}_0$.

2.3. Prior distribution for $\boldsymbol{\beta}$

The parameters β_x represent the age-specific pattern of mortality change: β_x indicates the sensitivity of the logarithm of the force of mortality at age x to variations in the time index. The shape of the β_x profile tells which rates decline rapidly and which slowly over time in response of change in κ_t.

The β_x profile is usually much more erratic (see Brouhns et al. (2002a) for an illustration with Belgian data). Some of the β_x's are close to zero (especially for the ages around the accident hump, for which mortality improvements are weak, as well as for older ages) while others are quite large (around birth for instance).

Our prior assumption for the β_x's is

$$\boldsymbol{\beta} \sim \mathcal{N}ormal_M(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_M).$$

In words, we start from the assumption that no mortality improvements occur for the population under study. The data will of course appropriately transform the prior distribution in case improvements do occur, as expected. Prior distributions for the hyperparameter σ_β^2 is taken to be inverse gamma, to facilitate the computation. Specifically,

$$\sigma_\beta^{-2} \sim \mathcal{G}amma(a_\beta, b_\beta)$$

for some constants a_β and b_β.

2.4. Prior distribution for $\boldsymbol{\alpha}$

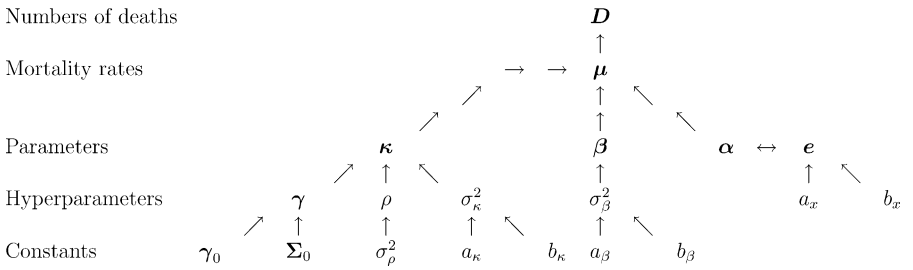
For technical reasons, it is more convenient to deal with the transformed vector $\boldsymbol{e} = \exp \boldsymbol{\alpha}$. The prior distribution for $\boldsymbol{e}$ is

$$e_x \sim \mathcal{G}amma(a_x, b_x)$$

for some constants a_x and b_x, with $x = x_{\min}, \ldots, x_{\max}$.

2.5. Summary of the model

The model can be summarized as follows:



The bottom line components are constants determining the distributions of the hyperparameters. The hyperparameters can be viewed on the line above. The third line represents the three sets of parameters of the Poisson log-bilinear model (1.4).

3. A posteriori distributions

3.1. MCMC approach

Henceforth, we use $f(\cdot)$ to denote a generic density function (the argument indicates to which random variables it pertains) and $f(\cdot|\cdot)$ to denote a generic conditional density function (where the conditioning variables follow the vertical bar).

Inference about the α_x's, β_x's and κ_t's is based on the posterior density of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa})$ given the mortality statistics, that is on the density

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}|\boldsymbol{D}) \propto f(\boldsymbol{D}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}). \tag{3.1}$$

We make inference empirically by collecting many realizations from the posterior distribution $f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}|\boldsymbol{D})$. This is done by simulation: it is possible to set up a Markov Chain whose stationary distribution is consistent with the posterior distribution (3.1). The Monte Carlo approach allows for inference based on sampling the posterior distribution of the parameters. A particularly attractive feature of this approach is the ease with which we can then explore the uncertainty associated with the estimates and the forecasts.

In this section, we discuss implementation of the Bayes procedures via Markov Chain Monte Carlo (MCMC). In particular, we use the Gibbs sampler and Metropolis–Hastings algorithm to generate samples from the posterior (3.1).

Let us now describe the MCMC procedure more formally. Consider a random vector $\boldsymbol{Z}$ with joint probability density function h. In a Bayesian context, some of the components of $\boldsymbol{Z}$ are model parameters, while others may represent unobserved past or future data. Suppose h is so complicated and analytically intractable that it does not permit independent random draws. In this case an MCMC simulation method may be used.

The main idea behind an MCMC method is to simulate realizations from a Markov Chain which has h as its stationary distribution. The resulting random draws $\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \ldots$ are no longer independent, but under mild regularity conditions (as described in the appendix of Smith and Roberts (1993), for example), the value of $\boldsymbol{Z}^{(k)}$ tends in distribution to that of a random draw from h as k becomes moderately large.

Determining how long an MCMC simulation should be run is a function of the particular application. Usually, several tens of thousands of iterations are enough. In any case, the first portion of the simulated Markov Chain is discarded in order to reduce the effect of the starting values. An ad hoc but useful test of convergence is obtained by running several simulations in parallel, with different starting values, and then comparing the results: the number of iterations must be increased if the results look rather different.

3.2. Metropolis–Hastings sampling for the time index vector $\boldsymbol{\kappa}$

Metropolis–Hastings algorithms produce Markov Chains whose stationary distribution is precisely (3.1) from which the samples have to be drawn. These algorithms are based on a Markov Chain whose dependence on the predecessor is split into two parts: a proposal and an acceptance of the proposal. The proposals suggest an arbitrary next step in the trajectory of the chain and the acceptance makes sure the appropriate limiting direction is maintained by rejecting unwanted moves of the chain.

Let us denote as

$$\boldsymbol{\kappa}_{-t} = (\kappa_{t_{\min}}, \ldots, \kappa_{t-1}, \kappa_{t+1}, \ldots, \kappa_{t_{\max}})'$$

the time index vector $\boldsymbol{\kappa}$ without its tth component. Denoting as

$$\boldsymbol{D}_t = (D_{x_{\min}t}, \ldots, D_{x_{\max}t})',$$

the vector of the D_{xt}'s, we define in the same way

$$\boldsymbol{D}_{-t} = (\boldsymbol{D}_{t_{\min}}, \ldots, \boldsymbol{D}_{t-1}, \boldsymbol{D}_{t+1}, \ldots, \boldsymbol{D}_{t_{\max}})'$$

as the matrix of the death counts D_{xt} without the column corresponding to calendar year t.

Now, each $\boldsymbol{\kappa}$ update is realized elementwise according to Metropolis–Hastings sampling. Specifically we look for the conditional probability density function $f(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$ that is, the density of κ_t given all other parameters and hyperparameters, as well as data points. Some manipulations yield

$$f(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) = \frac{f(\boldsymbol{\kappa}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)}{f(\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)}$$

$$= \frac{f(\kappa_{t_{\max}}, \boldsymbol{D}_{t_{\max}}|\boldsymbol{\kappa}_{-t_{\max}}, \boldsymbol{D}_{-t_{\max}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) f(\boldsymbol{\kappa}_{-t_{\max}}, \boldsymbol{D}_{-t_{\max}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)}{f(\boldsymbol{\kappa}_{-t}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)}.$$

Iterating this formula gives

$$f(\kappa_t|\boldsymbol{\kappa}_{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) = \frac{f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)}{f(\boldsymbol{\kappa}_{-t}, \boldsymbol{D}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)} f(\kappa_{t_{\min}}, \boldsymbol{D}_{t_{\min}}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\times \prod_{s=t_{\min}+1}^{t_{\max}} f(\kappa_s, \boldsymbol{D}_s|\kappa_{t_{\min}}, \ldots, \kappa_{s-1}, \boldsymbol{D}_{t_{\min}}, \ldots, \boldsymbol{D}_{s-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\propto f(\kappa_{t_{\min}}, \boldsymbol{D}_{t_{\min}}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\times \prod_{s=t_{\min}+1}^{t_{\max}} f(\kappa_s, \boldsymbol{D}_s|\kappa_{t_{\min}}, \ldots, \kappa_{s-1}, \boldsymbol{D}_{t_{\min}}, \ldots, \boldsymbol{D}_{s-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho).$$

Remember that the random vectors $\boldsymbol{D}_s$ are mutually independent given $\boldsymbol{\kappa}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho$. Moreover, their conditional distribution only depends on $(\kappa_s, \boldsymbol{\alpha}, \boldsymbol{\beta})$. This allows us to write

$$f(\kappa_s, \boldsymbol{D}_s|\kappa_{t_{\min}}, \ldots, \kappa_{s-1}, \boldsymbol{D}_{t_{\min}}, \ldots, \boldsymbol{D}_{s-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$= f(\boldsymbol{D}_s|\kappa_{t_{\min}}, \ldots, \kappa_{s-1}, \kappa_s, \boldsymbol{D}_{t_{\min}}, \ldots, \boldsymbol{D}_{s-1}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\times f(\kappa_s|\kappa_{t_{\min}},\ldots,\kappa_{s-1},\boldsymbol{D}_{t_{\min}},\ldots,\boldsymbol{D}_{s-1},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)$$
$$= f(\boldsymbol{D}_s|\kappa_s,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_s|\kappa_{t_{\min}},\ldots,\kappa_{s-1},\boldsymbol{D}_{t_{\min}},\ldots,\boldsymbol{D}_{s-1},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)$$
$$= f(\boldsymbol{D}_s|\kappa_s,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_s|\kappa_{s-1},\boldsymbol{\gamma},\sigma_\kappa^2,\rho).$$

Finally we find the following expression for the conditional density of κ_t:

$$f(\kappa_t|\boldsymbol{\kappa}_{-t},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)\propto f(\boldsymbol{D}_{t_{\min}}|\kappa_{t_{\min}},\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_{t_{\min}}|\boldsymbol{\gamma},\sigma_\kappa^2)\prod_{s=t_{\min}+1}^{t_{\max}} f(\boldsymbol{D}_s|\kappa_s,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_s|\kappa_{s-1},\boldsymbol{\gamma},\sigma_\kappa^2,\rho).$$

Let us now consider three cases:

(i) if $t = t_{\min}$ then

$$f(\kappa_t|\boldsymbol{\kappa}_{-t},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho) \propto f(\boldsymbol{D}_t|\kappa_t,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_t|\boldsymbol{\gamma},\sigma_\kappa^2)f(\kappa_{t+1}|\kappa_t,\boldsymbol{\gamma},\sigma_\kappa^2,\rho)$$
$$\propto \prod_x \exp(-E_{xt}\,\exp(\alpha_x+\beta_x\kappa_t))\prod_x \exp(\beta_x\kappa_t D_{xt})\,\exp\left(-\frac{1}{2\sigma_\kappa^2}(\kappa_t-\eta_t)^2\right)$$
$$\times\,\exp\left(-\frac{1}{2\sigma_\kappa^2}(\kappa_{t-1}-\eta_{t-1}-\rho(\kappa_t-\eta_t))^2\right); \tag{3.2}$$

(ii) if $t = t_{\max}$ then

$$f(\kappa_t|\boldsymbol{\kappa}_{-t},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho) \propto f(\boldsymbol{D}_t|\kappa_t,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_t|\kappa_{t-1},\boldsymbol{\gamma},\sigma_\kappa^2,\rho)$$
$$\propto \prod_x \exp(-E_{xt}\,\exp(\alpha_x+\beta_x\kappa_t))\prod_x \exp(\beta_x\kappa_t D_{xt})\,\exp\left(-\frac{1}{2\sigma_\kappa^2}(\kappa_t-\eta_t-\rho(\kappa_{t-1}-\eta_{t-1}))^2\right); \tag{3.3}$$

(iii) if $t_{\min} < t < t_{\max}$ then

$$f(\kappa_t|\boldsymbol{\kappa}_{-t},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{D},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho) \propto f(\boldsymbol{D}_t|\kappa_t,\boldsymbol{\alpha},\boldsymbol{\beta})f(\kappa_t|\kappa_{t-1},\boldsymbol{\gamma},\sigma_\kappa^2,\rho)f(\kappa_{t+1}|\kappa_t,\boldsymbol{\gamma},\sigma_\kappa^2,\rho)$$
$$\propto \prod_x \exp(-E_{xt}\,\exp(\alpha_x+\beta_x\kappa_t))\prod_x \exp(\beta_x\kappa_t D_{xt})\,\exp\left(-\frac{1}{2\sigma_\kappa^2}(\kappa_t-\eta_t-\rho(\kappa_{t-1}-\eta_{t-1}))^2\right)$$
$$\times\,\exp\left(-\frac{1}{2\sigma_\kappa^2}(\kappa_{t+1}-\eta_{t+1}-\rho(\kappa_t-\eta_t))^2\right). \tag{3.4}$$

Let us decompose iteration $(i+1)$ of Metropolis–Hastings sampling for κ_t update. Suppose that we have to update the value $\kappa_t^{(i)}$ obtained at iteration i. At this stage, we have already updated the preceding kappa's, so that we have at our disposal the $\kappa_s^{(i+1)}$ for $s < t$. We then proceed as follows:

1. First generate a candidate κ_t^* from the $\mathcal{N}ormal(\kappa_t^{(i)},\sigma_t^2)$ distribution with known variance σ_t^2.
2. Then compute acceptance probability

$$\psi(\kappa_t^{(i)},\kappa_t^*) = \min\left(1,\frac{f(\kappa_t^*|\boldsymbol{\kappa}_{-t}^{(i)},\boldsymbol{D},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)}{f(\kappa_t^{(i)}|\boldsymbol{\kappa}_{-t}^{(i)},\boldsymbol{D},\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)}\right),$$

where

$$\boldsymbol{\kappa}_{-t}^{(i)} = (\kappa_{t_{\min}}^{(i+1)}, \ldots, \kappa_{t-1}^{(i+1)}, \kappa_{t+1}^{(i)}, \ldots, \kappa_{t_{\max}}^{(i)})'.$$

3. Afterwards generate a realization u from the $Uniform(0, 1)$ distribution. If $u \leq \psi(\kappa_t^{(i)}, \kappa_t^*)$ then the candidate is kept and $\kappa_t^{(i+1)} = \kappa_t^*$. On the contrary, if $u > \psi(\kappa_t^{(i)}, \kappa_t^*)$ then the candidate is rejected and the Markov Chain does not move ($\kappa_t^{(i+1)} = \kappa_t^{(i)}$).
4. Finally we have to transform

$$\boldsymbol{\kappa}^{(i+1)} = (\kappa_{t_{\min}}^{(i+1)}, \ldots, \kappa_t^{(i+1)}, \kappa_{t+1}^{(i)}, \ldots, \kappa_{t_{\max}}^{(i)})'$$

and $\boldsymbol{\alpha}^{(i)}$ in order to fulfill the constraints (1.3). To this end, we use the following formulas:

$$\boldsymbol{\kappa}^{(i+1)} \leftarrow \boldsymbol{\kappa}^{(i+1)} - \overline{\kappa}, \qquad \boldsymbol{\alpha}^{(i)} \leftarrow \boldsymbol{\alpha}^{(i)} + \boldsymbol{\beta}^{(i)}\overline{\kappa},$$

where

$$\overline{\kappa} = \frac{1}{T}\left(\sum_{s \leq t} \kappa_s^{(i+1)} + \sum_{s > t} \kappa_s^{(i)}\right).$$

We remark that the choice of parameter σ_t^2 is free but not neglectable. It directly influences the acceptance rate of the proposals: a large variance will reduce the chance for the candidate to be kept and for the chain to move to another state. In practice we want the acceptation probability to be in the interval [20%, 50%]. Therefore, a trial and error method is used to select σ_t^2. Starting from some initial value, we compute the acceptation probability (on about one hundred iterations, say). If it is too small, we have to increase the variance σ_t^2 (make it double, say). On the contrary, if more than half the candidates are kept, we reduce the value of σ_t^2.

3.3. Metropolis–Hastings sampling for $\boldsymbol{\beta}$

The $\boldsymbol{\beta}$ update is quite similar to the $\boldsymbol{\kappa}$ one. Let us define

$$\boldsymbol{\beta}_{-x} = (\beta_{x_{\min}}, \ldots, \beta_{x-1}, \beta_{x+1}, \ldots, \beta_{x_{\max}})'$$

and in the same way

$$\boldsymbol{D}_{-x} = (\boldsymbol{D}_{x_{\min}}, \ldots, \boldsymbol{D}_{x-1}, \boldsymbol{D}_{x+1}, \ldots, \boldsymbol{D}_{x_{\max}})',$$

where $\boldsymbol{D}_x$ is the $(x - x_{\min} + 1)$th row of $\boldsymbol{D}$. With the same developments as in previous section we find

$$f(\beta_x|\boldsymbol{\beta}_{-x}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) \propto \prod_{y=x_{\min}}^{x_{\max}} f(\beta_y, \boldsymbol{D}_y|\beta_{x_{\min}}, \ldots, \beta_{y-1}, \boldsymbol{D}_{x_{\min}}, \ldots, \boldsymbol{D}_{y-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\propto \prod_{y=x_{\min}}^{x_{\max}} f(\boldsymbol{D}_y|\beta_{x_{\min}}, \ldots, \beta_{y-1}, \beta_y, \boldsymbol{D}_{x_{\min}}, \ldots, \boldsymbol{D}_{y-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho)$$

$$\times \prod_{y=x_{\min}}^{x_{\max}} f(\beta_y|\beta_{x_{\min}},\ldots,\beta_{y-1},\boldsymbol{D}_{x_{\min}},\ldots,\boldsymbol{D}_{y-1},\boldsymbol{\alpha},\boldsymbol{\kappa},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)$$

$$\propto \prod_{y=x_{\min}}^{x_{\max}} f(\boldsymbol{D}_y|\beta_y,\boldsymbol{\alpha},\boldsymbol{\kappa}) f(\beta_y) \propto f(\boldsymbol{D}_x|\beta_x,\boldsymbol{\alpha},\boldsymbol{\kappa}) f(\beta_x)$$

$$\propto \prod_t \exp(-E_{xt}\ \exp(\alpha_x+\beta_x\kappa_t)) \prod_t \exp(\beta_x\kappa_t D_{xt})\ \exp\left(-\frac{1}{2\sigma_\beta^2}\beta_x^2\right). \tag{3.5}$$

Now we can decompose iteration $(i+1)$ of Metropolis–Hastings sampling for β_x update. Suppose the parameter is estimated at iteration i by $\beta_x^{(i)}$ and we have estimations $\beta_y^{(i+1)}$ for $y<x$. We then proceed as follows:

1. Select a candidate β_x^* from the $\mathcal{N}ormal(\beta_x^{(i)},\sigma_x^2)$ distribution with known variance σ_x^2.
2. Compute the acceptance probability

$$\psi(\beta_x^{(i)},\beta_x^*) = \min\left(1, \frac{f(\beta_x^*|\boldsymbol{\beta}_{-x}^{(i)},\boldsymbol{D},\boldsymbol{\alpha},\boldsymbol{\kappa},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)}{f(\beta_x^{(i)}|\boldsymbol{\beta}_{-x}^{(i)},\boldsymbol{D},\boldsymbol{\alpha},\boldsymbol{\kappa},\sigma_\kappa^2,\sigma_\beta^2,\boldsymbol{\gamma},\rho)}\right),$$

   where

$$\boldsymbol{\beta}_{-x}^{(i)} = (\beta_{x_{\min}}^{(i+1)},\ldots,\beta_{x-1}^{(i+1)},\beta_{x+1}^{(i)},\ldots,\beta_{x_{\max}}^{(i)})'.$$

3. Afterwards generate a realization u from the $Uniform(0,1)$ distribution. If $u \le \psi(\beta_x^{(i)},\beta_x^*)$, the candidate is kept and $\beta_x^{(i+1)} = \beta_x^*$. On the contrary, if $u > \psi(\beta_x^{(i)},\beta_x^*)$, the candidate is rejected and the Markov Chain does not move $(\beta_x^{(i+1)} = \beta_x^{(i)})$.
4. Finally we have to transform vectors

$$\boldsymbol{\beta}^{(i+1)} = (\beta_{x_{\min}}^{(i+1)},\ldots,\beta_x^{(i+1)},\beta_{x+1}^{(i)},\ldots,\beta_{x_{\max}}^{(i)})'$$

   and $\boldsymbol{\kappa}^{(i+1)}$ in order to fulfill constraints (1.3):

$$\boldsymbol{\beta}^{(i+1)} \leftarrow \frac{\boldsymbol{\beta}^{(i+1)}}{\boldsymbol{\beta}_\bullet}, \qquad \boldsymbol{\kappa}^{(i+1)} \leftarrow \boldsymbol{\kappa}^{(i+1)}\boldsymbol{\beta}_\bullet,$$

   where

$$\boldsymbol{\beta}_\bullet = \sum_{y\le x} \beta_y^{(i+1)} + \sum_{y>x} \beta_y^{(i)}.$$

As for the $\boldsymbol{\kappa}$ update, the variance σ_x^2 must be adjusted in order to have acceptance probabilities between 20 and 50%.

3.4. Gibbs sampling for $\boldsymbol{\alpha}$

Another standard approach which produces a Markov Chain whose stationary distribution is consistent with the posterior distribution (3.1) is based on a variant of the Metropolis algorithm called the Gibbs sampler. It will enable us to exploit conditional densities to obtain realizations from the posterior density. The Gibbs sampler requires that the unknown model parameters are first assigned arbitrary values. Then an iterative sampling process takes place. At each iteration, the Gibbs sampler visits each unknown parameter in turn, and generates a random value from its full conditional distribution, conditional upon current values of all other parameters and upon the data. Then, each iteration yields a sample realization of the complete set of unknown parameters in the model. The generated realizations converge in distribution to the joint posterior distribution of the unknown parameters.

Let us consider the likelihood function (2.1) as function of the vector $\boldsymbol{e}$ only:

$$L(\boldsymbol{e}, \boldsymbol{\beta}, \boldsymbol{\kappa}) \propto \prod_x \prod_t \exp(-E_{xt}\ \exp(\alpha_x + \beta_x\kappa_t) + D_{xt}(\alpha_x + \beta_x\kappa_t)) \propto \prod_x \exp(-c_x e_x) e_x^{D_{x\bullet}}$$

with

$$c_x = \sum_t E_{xt}\ \exp(\beta_x\kappa_t) \quad \text{and} \quad D_{x\bullet} = \sum_t D_{xt}.$$

To draw random samples from the posterior density, we use the Gibbs sampling algorithm. The essence of the Gibbs sampler lies in breaking a complicated joint probability density into a set of full conditional densities, and sampling one variable at a time, conditional on the values of the others.

For $x = x_{\min}, \ldots, x_{\max}$, we can write

$$f(e_x|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) = f(e_x|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}) \propto f(e_x, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}) f(e_x) \propto \exp(-c_x e_x) e_x^{D_{x\bullet}} e_x^{a_x - 1}\ \exp(-b_x e_x)$$
$$\propto \exp(-(b_x + c_x)e_x) e_x^{a_x + D_{x\bullet} - 1}$$

so that the distribution of e_x given $\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho$ is still gamma with updated parameters, that is,

$$(e_x|\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) \sim \mathcal{G}amma(a_x + D_{x\bullet}, b_x + c_x). \tag{3.6}$$

Realizations of e_x given $\boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}, \rho$ are thus easily generated.

3.5. Gibbs sampling for ρ

From conditional distribution definition we have

$$f(\rho|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) = \frac{f(\rho, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma})}{f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma})} \propto f(\boldsymbol{D}|\rho, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) f(\rho, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma})$$
$$= f(\boldsymbol{D}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) f(\boldsymbol{\kappa}, \rho|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) \propto f(\boldsymbol{\kappa}, \rho) = f(\boldsymbol{\kappa}|\rho) f(\rho).$$

The conditional density of $\boldsymbol{\kappa}$ given ρ is given by

$$f(\boldsymbol{\kappa}|\rho) = \prod_t f(\kappa_t|\kappa_{t-1}, \rho) \propto \exp\left(-\frac{1}{2\sigma_\kappa^2}(a_\rho \rho^2 - 2b_\rho \rho)\right)$$

with

$$a_\rho = \sum_t (\kappa_{t-1} - \eta_{t-1})^2 \quad \text{and} \quad b_\rho = \sum_t (\kappa_t - \eta_t)(\kappa_{t-1} - \eta_{t-1}),$$

where the notation $\eta_t = \gamma_1 + \gamma_2 t$ introduced in Section 2.2 has been used and with the convention $\kappa_{t_{\min}-1} = \eta_{t_{\min}-1}$. Therefore,

$$f(\rho|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) \propto \exp\left(-\frac{1}{2\sigma_\kappa^2}(a_\rho \rho^2 - 2b_\rho \rho)\right) \exp\left(-\frac{1}{2\sigma_\rho^2}\rho^2\right) \propto \exp\left(-\frac{1}{2\sigma_\rho^{2*}}(\rho - \mu_\rho^*)^2\right)$$

with

$$\mu_\rho^* = \frac{b_\rho}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}} \quad \text{and} \quad \sigma_\rho^{2*} = \frac{\sigma_\kappa^2}{a_\rho + \frac{\sigma_\kappa^2}{\sigma_\rho^2}}.$$

The distribution of ρ given $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}$ can then be written as

$$(\rho|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \sigma_\beta^2, \boldsymbol{\gamma}) \sim \mathcal{N}ormal(\mu_\rho^*, \sigma_\rho^{2*}) \text{ truncated to } (-1, 1). \tag{3.7}$$

Simulation from (3.7) is easy.

3.6. Gibbs sampling for σ_κ^2

By using the same principles as previously we find

$$f(\sigma_\kappa^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) \propto f(\boldsymbol{\kappa}|\sigma_\kappa^2) f(\sigma_\kappa^2).$$

Since

$$f(\boldsymbol{\kappa}|\sigma_\kappa^2) = \frac{1}{2\pi\sigma_\kappa^2} \exp\left(-\frac{1}{2\sigma_\kappa^2}\sum_t (\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right)$$

and the prior distribution of σ_κ^{-2} is $\mathcal{G}amma(a_\kappa, b_\kappa)$, we find

$$f(\sigma_\kappa^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) \propto {\sigma_\kappa}^{-2(a_\kappa+1+T/2)} \exp\left(-\frac{1}{\sigma_\kappa^2}\left(b_\kappa + \frac{1}{2}\sum_t (\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right)\right)$$

with $\kappa_{t_{\min}-1} = \eta_{t_{\min}-1}$. The distribution of σ_κ^{-2} given $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\beta^2, \boldsymbol{\gamma}, \rho$ is then given by

$$(\sigma_\kappa^{-2}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\beta^2, \boldsymbol{\gamma}, \rho) \sim \mathcal{G}amma\left(a_\kappa + \frac{T}{2}, b_\kappa + \frac{1}{2}\sum_t (\kappa_t - \eta_t - \rho(\kappa_{t-1} - \eta_{t-1}))^2\right). \tag{3.8}$$

3.7. Gibbs sampling for σ_β^2

Applying the same reasoning as before, we find

$$f(\sigma_\beta^2|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \boldsymbol{\gamma}, \rho) \propto f(\boldsymbol{\beta}|\sigma_\beta^2) f(\sigma_\beta^2) \propto \sigma_\beta^{-2(a_\beta+1+M/2)} \exp\left(-\frac{1}{\sigma_\beta^2}\left(b_\beta + \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right).$$

The distribution of σ_β^{-2} given $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, $\boldsymbol{D}$, σ_κ^2, $\boldsymbol{\gamma}$, ρ is then given by

$$(\sigma_\beta^{-2}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{D}, \sigma_\kappa^2, \boldsymbol{\gamma}, \rho) \sim \mathcal{G}amma\left(a_\beta + \frac{M}{2}, b_\beta + \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\beta}\right). \tag{3.9}$$

3.8. Gibbs sampling for $\boldsymbol{\gamma}$

As previously we have

$$f(\boldsymbol{\gamma}|\boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \rho) \propto f(\boldsymbol{\kappa}|\boldsymbol{\gamma}, \rho, \sigma_\kappa^2) f(\boldsymbol{\gamma}).$$

Because $\boldsymbol{\gamma}$ is a priori distributed according to $\mathcal{N}ormal_2(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0)$, we find

$$\begin{aligned} f(\boldsymbol{\gamma}|\boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \rho) &\propto \exp\left(-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\kappa} - \boldsymbol{X}\boldsymbol{\gamma})' \boldsymbol{Q}(\boldsymbol{\kappa} - \boldsymbol{X}\boldsymbol{\gamma})\right) \exp\left(-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_\kappa^2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)\right) \end{aligned}$$

with

$$\boldsymbol{\Sigma}^* = (\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1})^{-1} \quad \text{and} \quad \boldsymbol{\gamma}^* = \boldsymbol{\Sigma}^*(\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{\kappa} + \sigma_\kappa^2\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0).$$

Thus the distribution of $\boldsymbol{\gamma}$ given $\boldsymbol{\kappa}$, σ_κ^2, σ_β^2, ρ is

$$(\boldsymbol{\gamma}|\boldsymbol{\kappa}, \sigma_\kappa^2, \sigma_\beta^2, \rho) \sim \mathcal{N}ormal_2(\boldsymbol{\gamma}^*, \sigma_\kappa^2\boldsymbol{\Sigma}^*). \tag{3.10}$$

4. Numerical illustration

4.1. Data set

Our data are about French male population aged 0 to 89 between 1950 and 2000. The data related to calendar years 1950–1997 come from INED (*Institut National d'Etudes Démographiques* based in Paris, France). Those related to calendar years 1998, 1999 and 2000 have been obtained from INSEE (*Institut National de la Statistique et des Etudes Economiques* based in Paris, France). The following information is available: the numbers L_{xt} of people aged x on 1 January of year t (for x between 0 and 89), and the numbers D_{xt} of people aged x dying during year t. The exposure-to-risk E_{xt} is then computed under assumption (1.1).

4.2. Initialization and choice of prior distributions for the hyperparameters

As seen in the previous section, we work with three vectors of parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$) and with five hyperparameters (ρ, σ_κ^2, σ_β^2 and the two components of $\boldsymbol{\gamma}$). We also have to fix the constants a_κ, b_κ, a_β, b_β, σ_ρ^2, the M components a_x, the M components b_x, the vector $\boldsymbol{\gamma}_0$ and the matrix $\boldsymbol{\Sigma}_0$ involved in the distribution of the hyperparameters.

The choices for the constants determining the distributions of the hyperparameters are made in an empirical Bayes approach. In empirical Bayes, hyperparameters from the last level of a hierarchical model are estimated rather than chosen a priori. Although this procedure might seem better because it lets the data decide about reasonable values for obscure hyperparameters, many theoretical arguments have been levelled against it. Despite the inferential problems, this procedure is often used (namely because using the data in this way turns out to be equivalent to making the prior indifferent to certain chosen parameters; see Carlin and Louis (2000)).

Specifically, we first compute the frequentist estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ by maximizing the likelihood (2.1), as described in Brouhns et al. (2002a,b). Because of the presence of the bilinear term $\beta_x\kappa_t$, it is not possible to estimate the proposed model with commercial statistical packages that implement Poisson regression. A uni-dimensional or elementary Newton method is used instead as proposed by Goodman (1979) for estimating log-linear models with bilinear terms. The estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ obtained in this way are further referred to as Goodman estimates.

We take for $\boldsymbol{\gamma}_0$ the estimated parameters of a linear regression of $\boldsymbol{\kappa}$ on calendar time. Then, $\boldsymbol{\Sigma}_0$ is taken to be the estimated covariance matrix of these estimates. Afterwards, ρ and σ_κ^2 can be initialized by fitting an AR(1) model on $(\boldsymbol{\kappa} - \boldsymbol{\gamma}_0 X)$. Finally σ_β^2 is initialized to be the empirical variance of the Goodman $\widehat{\beta}_x$'s.

When the model (1.2) is fitted by ordinary least-squares, the fitted values of α_x exactly equal the average of $\ln \widehat{\mu_x}(t)$ over time t so that $\exp \alpha_x$ is the general shape of the mortality schedule. Also, in the Poisson log-bilinear model, the fitted $\exp \alpha_x$ mimick the observed average of the $\widehat{\mu_x}(t)$'s. In order to obtain an uninformative prior distribution (i.e. with large variance), we have to choose a small b_x, 0.001 say. Afterwards, a_x can be chosen equal to $b_x \exp \widehat{\alpha}_x$ (so that $\mathbb{E}[e_x] = \exp \widehat{\alpha}_x$).

In the same way, if $\sigma_\beta^{-2} \sim \mathcal{G}amma(a_\beta, b_\beta)$, then $\mathbb{E}[\sigma_\beta^2] = \frac{b_\beta}{a_\beta - 1}$ for $a_\beta > 1$ and $\mathbb{V}\mathrm{ar}[\sigma_\beta^2] = \frac{b_\beta^2}{(a_\beta-1)^2(a_\beta-2)}$ for $a_\beta > 2$. So constants a_β and b_β control prior mean and variance of σ_β^2. Taking a_β near to (but greater than) 2 will give a huge variance. So we find $b_\beta = (a_\beta - 1)\sigma_\beta^2$. The same argument can be used for a_κ and b_κ. We use $a_\beta = a_\kappa = 2.1$ for the application.

Finally the initial value of σ_ρ^2 does not seem to be very influential since ρ is restricted. We choose $\sigma_\rho^2 = 1$.

It is worth mentioning that other constants have also been used. Typically, we have increased the a priori variance of parameters and hyperparameters. These more diffuse prior choices were based on smaller values of b_x (from 0.001 to 0.00001), values of a_β closer to 2 (from 2.1 to 2.00001), and "greater" matrix $\boldsymbol{\Sigma}_0$ (up to 10 times the initial matrix). The results were similar to those obtained with the initial values deduced from the empirical Bayes approach.

Instead of basing the prior choices of the parameters on the Goodman's estimations of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$, we also used the Lee–Carter estimates to fix the constants, the hyperparameters and the parameters involved in the model (in an empirical Bayes setting). Again, the results obtained in this way were almost identical to those obtained with the Goodman's estimations.

4.3. Convergence diagnostics

In practice, we typically run the hybrid Metropolis–Hastings and Gibbs sampler for an initial period of a few thousands cycles and then collect information from several further thousands of cycles (of which we store every 10th for the subsequent construction of approximate interval estimates). The posterior means are estimated by the corresponding sample means. Here, 20,000 iterations are computed. The first 10,000 iterations are considered as the burn-in period. The last 10,000 iterations are used for estimation of the posterior distribution.

A sample from the distribution of interest (the posterior distribution of the parameter in our example) is only attained with MCMC when the number of iterations of the chain approaches infinity. This is of course impossible in practice and a value obtained at a sufficiently large iteration is taken instead. This raises the question of how large this iteration should be.

Some informal checks of convergence based on graphical techniques are commonly used. Several chains can be run in parallel with different initial states or a single chain is screened for exhibiting the same qualitative behavior through iterations (often after a transient initial period). More formal diagnostics are developed in Cowles and Carlin (1996) and Brooks and Roberts (1998).

Fig. 1 gives the selected σ_t^2 and σ_x^2 and corresponding acceptation probabilities. These variances are selected by a trial and error method. The algorithm starts with $\sigma_x^2 = \sigma_t^2 = 1$ for all x and t, and a first 100-iteration-pilot run is computed. Variance σ_x^2 (or σ_t^2) is doubled if the corresponding acceptation probability is below 20%; it is divided by

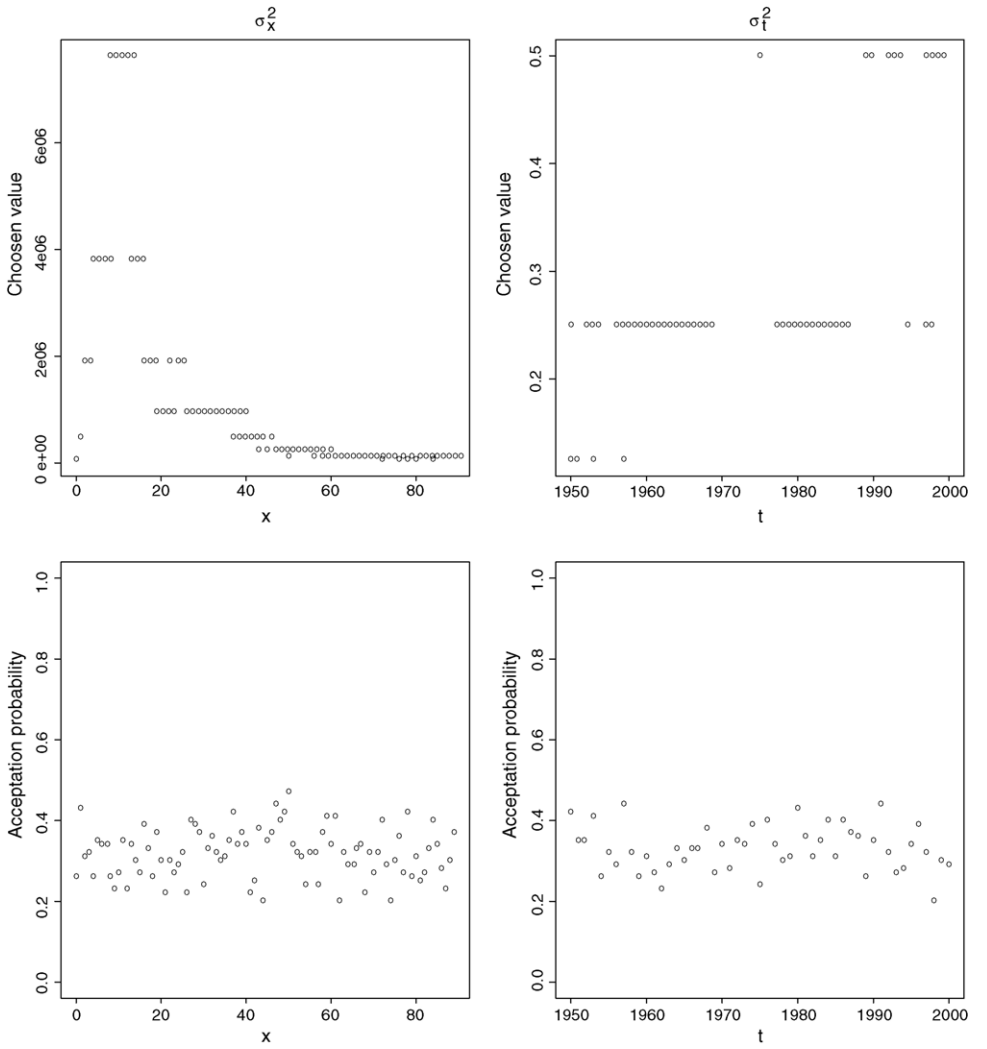


Fig. 1. Metropolis–Hastings sampling for β_x (to the left) and for κ_t (to the right). Chosen proposal variances (top panel with σ_x^2 to the left and σ_t^2 to the right) and corresponding acceptance rates after the last 100-iteration-pilot run (bottom panel).

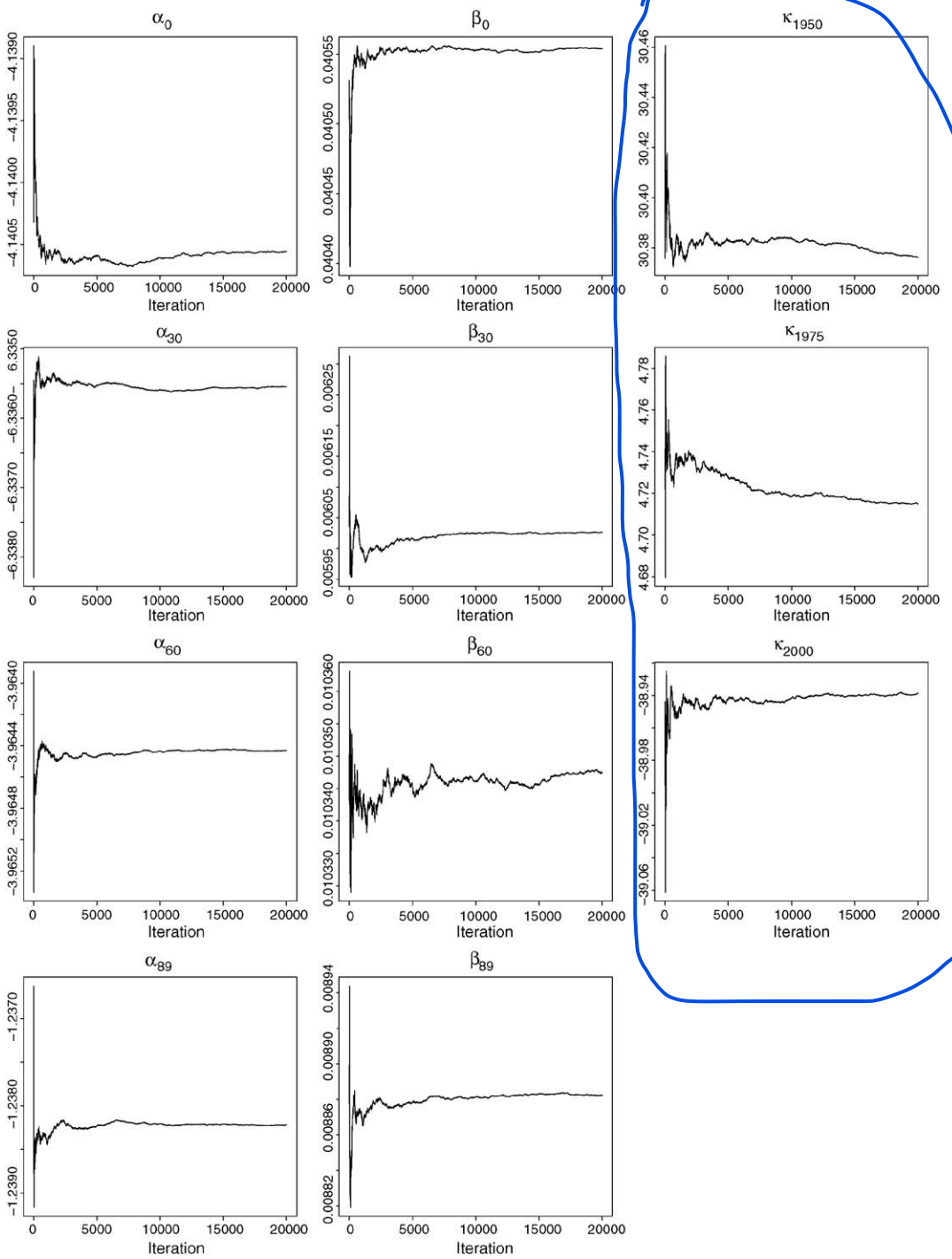


Fig. 2. Ergodic mean of parameters α_x (left), β_x (center) and κ_t (right) for some selected ages (0, 30, 60 and 89) and years (1950, 1975 and 2000).

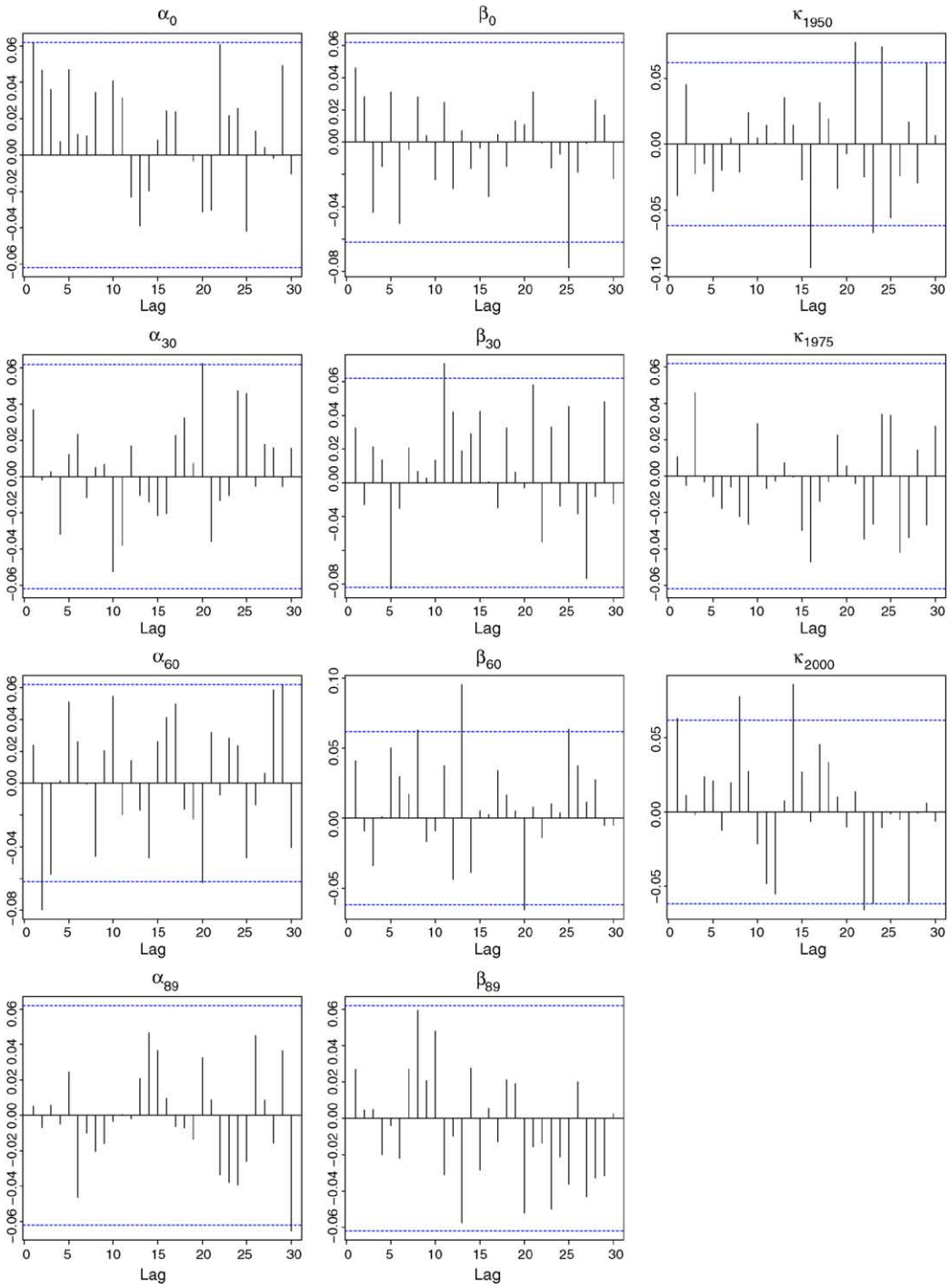


Fig. 3. Autocorrelations of parameters α_x (left), β_x (center) and κ_t (right) for some selected ages (0, 30, 60 and 89) and years (1950, 1975 and 2000) based on 1000 recorded iterations after a burn-in period of 10,000 iterations (lag 0 is not shown).

Table 1
Estimations of α_x, β_x and κ_t for several ages x and years t

α_x	x			
	0	30	60	89
MCMC mean	−4.14050	−6.33551	−3.96443	−1.23823
MCMC median	−4.14053	−6.33551	−3.96442	−1.23823
95% MCMC CI	[−4.14448,−4.13660]	[−6.34633,−6.32495]	[−3.96848,−3.96029]	[−1.24341,−1.23309]
Goodman estimate	−4.14004	−6.33545	−3.96447	−1.23822
Lee–Carter estimate	−4.11565	−6.34088	−3.96864	−1.23515

β_x	x			
	0	30	60	89
MCMC mean	0.04056	0.00603	0.01035	0.00888
MCMC median	0.04056	0.00603	0.01035	0.00888
95% MCMC CI	[0.04024,0.04086]	[0.00549,0.00658]	[0.01013,0.01057]	[0.00864,0.00913]
Goodman estimate	0.04050	0.00604	0.01034	0.00888
Lee–Carter estimate	0.03569	0.00656	0.01031	0.00924

κ_t	t		
	1950	1975	2000
MCMC mean	30.36856	4.71150	−38.93598
MCMC median	30.36549	4.71467	−38.93671
95% MCMC CI	[30.07773,30.66069]	[4.40581,5.01558]	[−39.35740,−38.52124]
Goodman estimate	30.38238	4.72366	−38.97546
Lee–Carter estimate	28.68970	6.75446	−40.10878

um pouco + alto

2 if corresponding acceptation probability is greater than 50%. A second 100-iteration-pilot run is then computed and variances σ_x^2 and σ_t^2 are adjusted as previously. The algorithm stops when rates for each age x and each year t are between 20 and 50%. We observe larger variances σ_x^2 where mortality is very specific, i.e. between ages 0 and 40 (new borns, accidental hump). In the same way variance σ_t^2 grows with year t.

Once we know how to sample from the a posteriori distribution of the parameters, we can compute a posteriori means by averaging the values obtained by repeating these steps a large number of times (after having discarded a suitable number of "burn-in" iterations to ensure that the algorithm has converged). Note that we do not worry about autocorrelation as long as we are not computing standard errors. Fig. 2 shows the trajectory of the ergodic averages for all the three parameters at select ages or years. Clearly, the asymptotic behavior is attained after the first 10,000 runs.

In order to get samples with no significant autocorrelations, we take only one realization of each parameter by each 10-iteration-block after a burn-in period of 10,000 iterations. We then have a sample of size 10,000/10 = 1000 from the a posteriori distribution of each parameter. Autocorrelations of these series are depicted in Fig. 3, clearly showing that no significant autocorrelations remain in the samples (recall that if the autocorrelation falls outside the interval determined by the two horizontal dotted lines then it significantly differs from 0).

thin = 10

4.4. Comparison with other methods

Since the Lee–Carter method is the most popular one, it is interesting to compare MCMC estimates to this benchmark. To obtain the Lee–Carter estimates the model (1.2) is fitted to the matrix of age-specific observed forces of mortality using singular value decomposition (SVD). Then, $\widehat{\alpha_x}$ is taken to be the row average of the ln $\widehat{\mu_x}(t)$'s, and the $\widehat{\beta_x}$'s and $\widehat{\kappa_t}$'s are obtained from the first term of an SVD of the matrix ln $\widehat{\mu_x}(t) - \widehat{\alpha_x}$.

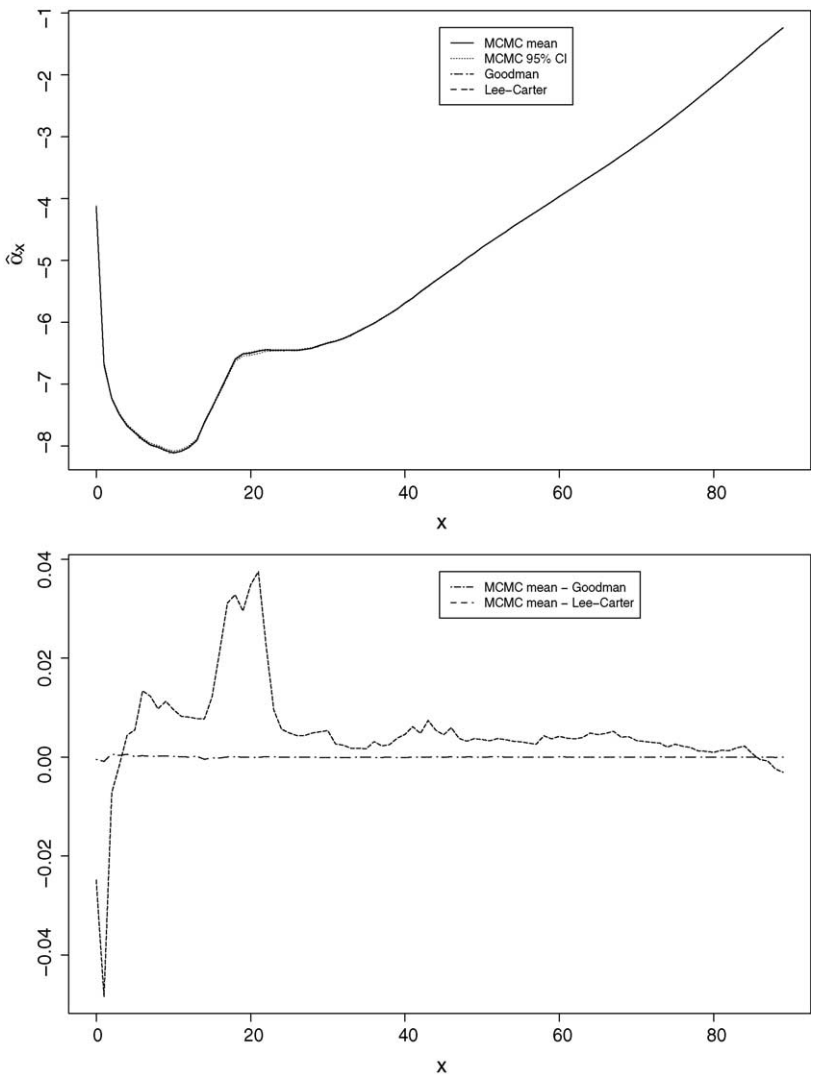


Fig. 4. Estimations of $\boldsymbol{\alpha}$, together with differences between Bayesian estimates and their Poisson Goodman and Lee–Carter counterparts.

Figs. 4–6 depict estimates of α_x's, β_x's and κ_t's according to the Lee–Carter SVD algorithm, the Poisson Goodman algorithm and the MCMC algorithm as well as the difference between the MCMC means and the corresponding Goodman and Lee–Carter estimates. Some numerical values for several ages x and years t can be viewed in Table 1. From Fig. 4, we see that the pattern of the estimated α_x is almost identical whatever the estimation method used. Some discrepancies between the Poisson and Lee–Carter estimates are visible for ages up to 30. On the contrary, Figs. 5 and 6 show that there are significant differences in the $\widehat{\kappa}_t$ and $\widehat{\beta}_x$ patterns. The values obtained with Lee–Carter SVD lie outside the MCMC credible intervals for numerous ages and calendar years. The largest differences appear for the κ_t's, for which the Lee–Carter approach produces smaller values from 1960 to 1990, and larger values after 1990. This will affect the mortality projections.

Table 2 displays the mortality rates for the generation aged 39 in calendar year 1950. Whereas the values obtained with the Poisson log-bilinear model always lie inside the 95% MCMC credible interval, those derived from the

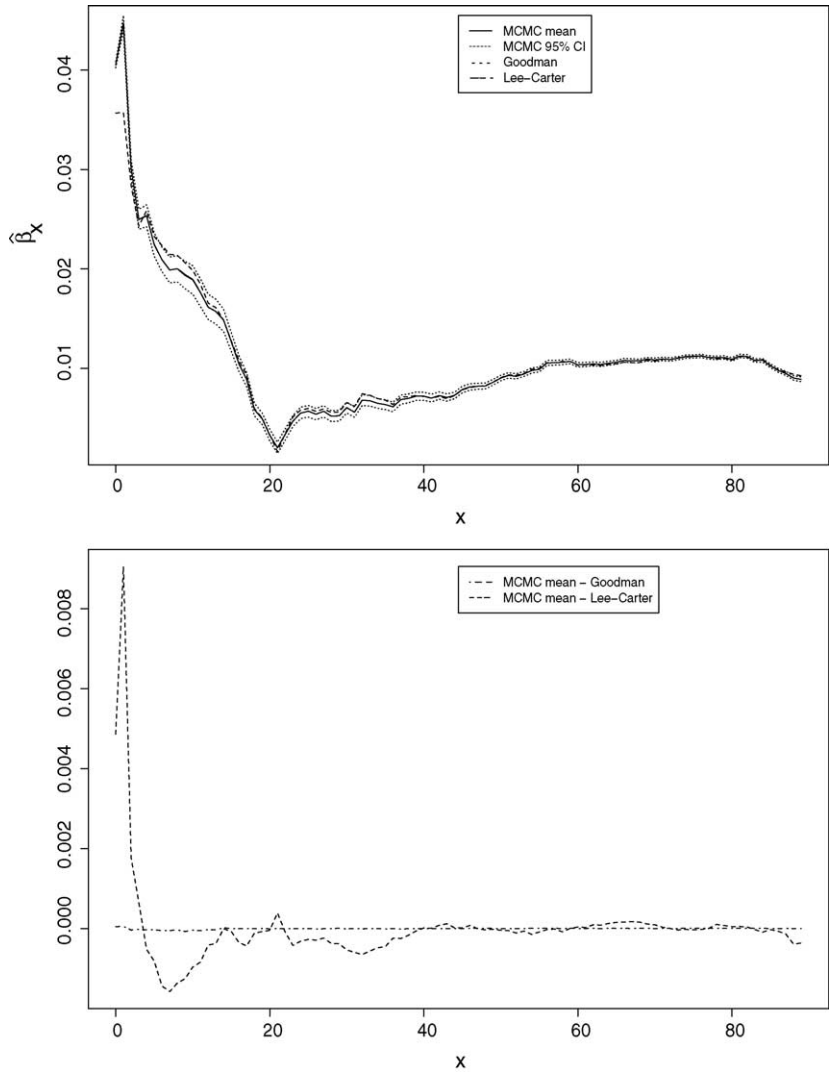


Fig. 5. Estimations of $\boldsymbol{\beta}$, together with differences between Bayesian estimates and their Poisson Goodman and Lee–Carter counterparts.

Lee–Carter methodology are about half the time outside this interval. These results highlight significant differences with Lee–Carter estimates. They are nevertheless in agreement with those obtained with the Poisson log-bilinear model. Note also that the quality of the fit is rather poor whatever the method considered. This is usually the case for this kind of "robust" model used for projection purposes. Measuring the quality of the fit for complex hierarchical Bayesian models is somewhat difficult. A Bayesian measure of model complexity and fit has been proposed by Spiegelhalter et al. (2002). Invoking an information theoretic argument, they derived a measure for the effective number of parameters in a model as the difference between the posterior mean of the deviance and the deviance at the posterior means of the parameters of interest. This yields a deviance information criterion (abbreviated as DIC) for measuring the goodness-of-fit and for comparing models. Such a criterion has been used in Denuit and Lang (2004) for comparing different regression models for claim frequencies and claim costs in nonlife insurance.

Table 2
Mortality rates for generation aged 39 in 1950

Age	Year	Empirical	MCMC mean	MCMC median	MCMC 95% CI	Goodman	Lee–Carter
39	1950	0.00416	0.00383	0.00383	[0.00377,0.00389]	0.00383	0.00377
40	1951	0.00450	0.00425	0.00425	[0.00418,0.00431]	0.00425	0.00424
41	1952	0.00432	0.00444	0.00444	[0.00439,0.00450]	0.00444	0.00436
42	1953	0.00474	0.00498	0.00498	[0.00492,0.00505]	0.00498	0.00499
43	1954	0.00509	0.00530	0.00530	[0.00524,0.00537]	0.00530	0.00520
44	1955	0.00556	0.00581	0.00581	[0.00575,0.00588]	0.00581	0.00576
45	1956	0.00642	0.00646	0.00646	[0.00639,0.00653]	0.00646	0.00649
46	1957	0.00648	0.00700	0.00700	[0.00693,0.00707]	0.00700	0.00698
47	1958	0.00613	0.00728	0.00729	[0.00723,0.00735]	0.00728	0.00716
48	1959	0.00756	0.00805	0.00805	[0.00798,0.00811]	0.00805	0.00795
49	1960	0.00810	0.00875	0.00875	[0.00868,0.00882]	0.00875	0.00873
50	1961	0.00848	0.00946	0.00946	[0.00940,0.00953]	0.00946	0.00936
51	1962	0.00992	0.01058	0.01058	[0.01050,0.01065]	0.01058	0.01064
52	1963	0.01071	0.01156	0.01156	[0.01148,0.01164]	0.01157	0.01168
53	1964	0.01141	0.01200	0.01200	[0.01192,0.01207]	0.01199	0.01195
54	1965	0.01251	0.01332	0.01332	[0.01323,0.01340]	0.01332	0.01352
55	1966	0.01336	0.01415	0.01414	[0.01406,0.01423]	0.01414	0.01418
56	1967	0.01483	0.01547	0.01547	[0.01538,0.01556]	0.01547	0.01566
57	1968	0.01601	0.01682	0.01682	[0.01673,0.01692]	0.01682	0.01706
58	1969	0.01804	0.01859	0.01859	[0.01849,0.01870]	0.01860	0.01905
59	1970	0.01796	0.01908	0.01908	[0.01898,0.01919]	0.01908	0.01919
60	1971	0.01967	0.02076	0.02076	[0.02066,0.02087]	0.02077	0.02097
61	1972	0.02152	0.02209	0.02209	[0.02198,0.02221]	0.02209	0.02239
62	1973	0.02294	0.02377	0.02377	[0.02366,0.02389]	0.02377	0.02411
63	1974	0.02491	0.02545	0.02545	[0.02533,0.02557]	0.02545	0.02576
64	1975	0.02658	0.02753	0.02753	[0.02740,0.02767]	0.02753	0.02797
65	1976	0.02877	0.02950	0.02950	[0.02936,0.02965]	0.02951	0.02996
66	1977	0.02942	0.03092	0.03092	[0.03077,0.03107]	0.03091	0.03110
67	1978	0.03269	0.03351	0.03351	[0.03335,0.03367]	0.03351	0.03384
68	1979	0.03511	0.03584	0.03584	[0.03567,0.03601]	0.03584	0.03619
69	1980	0.03723	0.03864	0.03864	[0.03846,0.03882]	0.03864	0.03888
70	1981	0.03988	0.04192	0.04192	[0.04173,0.04213]	0.04193	0.04227
71	1982	0.04222	0.04431	0.04431	[0.04410,0.04453]	0.04430	0.04455
72	1983	0.04673	0.04859	0.04860	[0.04836,0.04882]	0.04861	0.04907
73	1984	0.04927	0.05161	0.05161	[0.05136,0.05186]	0.05159	0.05187
74	1985	0.05456	0.05660	0.05660	[0.05632,0.05688]	0.05661	0.05688
75	1986	0.05851	0.06087	0.06087	[0.06055,0.06118]	0.06088	0.06099
76	1987	0.06167	0.06416	0.06416	[0.06381,0.06451]	0.06415	0.06387
77	1988	0.06581	0.06909	0.06910	[0.06870,0.06948]	0.06908	0.06860
78	1989	0.07149	0.07547	0.07547	[0.07504,0.07589]	0.07547	0.07526
79	1990	0.07539	0.08126	0.08126	[0.08079,0.08175]	0.08126	0.08090
80	1991	0.08408	0.08879	0.08879	[0.08825,0.08931]	0.08879	0.08829
81	1992	0.08993	0.09473	0.09473	[0.09411,0.09534]	0.09472	0.09418
82	1993	0.10273	0.10490	0.10490	[0.10424,0.10557]	0.10493	0.10452
83	1994	0.11018	0.11329	0.11329	[0.11253,0.11404]	0.11326	0.11254
84	1995	0.12268	0.12450	0.12449	[0.12364,0.12538]	0.12451	0.12383
85	1996	0.13762	0.13943	0.13942	[0.13847,0.14039]	0.13943	0.13873
86	1997	0.14778	0.15394	0.15394	[0.15280,0.15505]	0.15390	0.15263
87	1998	0.16644	0.16937	0.16937	[0.16802,0.17067]	0.16938	0.16753
88	1999	0.19270	0.18918	0.18919	[0.18760,0.19070]	0.18920	0.18570
89	2000	0.20178	0.20513	0.20513	[0.20325,0.20701]	0.20511	0.20071

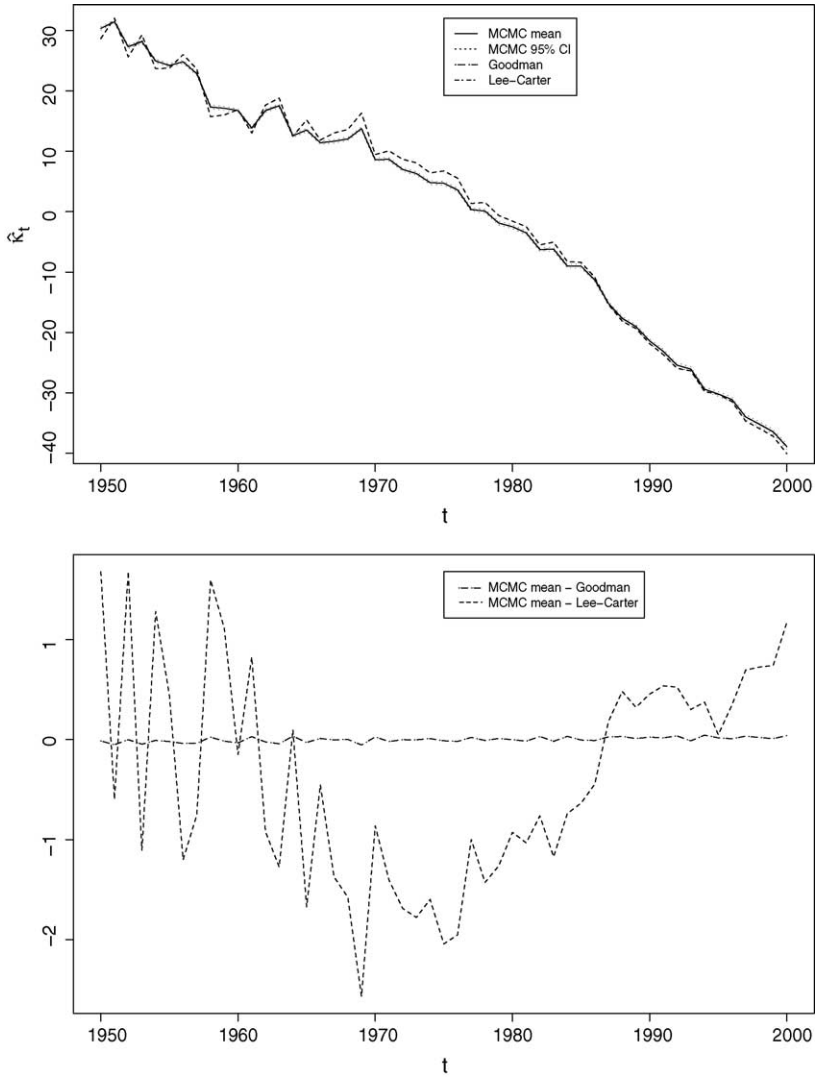


Fig. 6. Estimations of κ, together with differences between Bayesian estimates and their Poisson Goodman and Lee–Carter counterparts.

We mention that the close agreement between the Poisson Goodman estimates and the results obtained from a Bayesian model based on the same Poisson likelihood (especially if Poisson Goodman empirical Bayes prior estimates are used) was expected. Similarly, a Bayesian model based on the Gaussian likelihood would have lead to results comparable to Lee–Carter estimates. Note that since the data consist of counts, a Poisson formulation is more natural than a Gaussian one.

Let us now examine the forecasts obtained from the different models. To this end, we first project the κ_t's to the future (i.e. beyond year 2000) with the different models. This is shown in Fig. 7. The projection for the frequentist Poisson and the Lee–Carter models has been performed with the help of a random walk with drift. In the Bayesian approach, realizations of the future κ_t's are generated as part of the MCMC algorithm (this automatically provides credible intervals). We can see that all the projected κ_t's lie inside the 95% MCMC credible intervals (whatever the

Table 3
Projected mortality rates for generation aged 39 in 2000

Age	Year	MCMC mean	MCMC median	MCMC 95% CI	Goodman	Lee–Carter
39	2000	0.00232	0.00232	[0.00228,0.00237]	0.00232	0.00230
40	2001	0.00254	0.00254	[0.00246,0.00262]	0.00253	0.00250
41	2002	0.00276	0.00276	[0.00265,0.00286]	0.00274	0.00270
42	2003	0.00300	0.00300	[0.00286,0.00314]	0.00297	0.00294
43	2004	0.00328	0.00328	[0.00311,0.00345]	0.00325	0.00322
44	2005	0.00353	0.00353	[0.00333,0.00373]	0.00349	0.00345
45	2006	0.00372	0.00372	[0.00348,0.00396]	0.00366	0.00362
46	2007	0.00398	0.00398	[0.00369,0.00425]	0.00391	0.00387
47	2008	0.00427	0.00428	[0.00394,0.00459]	0.00419	0.00414
48	2009	0.00467	0.00467	[0.00428,0.00503]	0.00458	0.00452
49	2010	0.00492	0.00492	[0.00447,0.00533]	0.00481	0.00475
50	2011	0.00525	0.00526	[0.00474,0.00572]	0.00512	0.00504
51	2012	0.00555	0.00556	[0.00498,0.00607]	0.00540	0.00532
52	2013	0.00599	0.00600	[0.00534,0.00657]	0.00583	0.00572
53	2014	0.00632	0.00634	[0.00561,0.00696]	0.00614	0.00604
54	2015	0.00671	0.00672	[0.00590,0.00742]	0.00651	0.00637
55	2016	0.00712	0.00714	[0.00623,0.00792]	0.00690	0.00678
56	2017	0.00733	0.00735	[0.00634,0.00823]	0.00708	0.00698
57	2018	0.00781	0.00783	[0.00672,0.00878]	0.00754	0.00744
58	2019	0.00833	0.00836	[0.00715,0.00939]	0.00804	0.00788
59	2020	0.00885	0.00887	[0.00756,0.01002]	0.00854	0.00842
60	2021	0.00973	0.00977	[0.00833,0.01098]	0.00938	0.00928
61	2022	0.01036	0.01039	[0.00887,0.01173]	0.00998	0.00987
62	2023	0.01103	0.01106	[0.00939,0.01254]	0.01062	0.01055
63	2024	0.01188	0.01192	[0.01012,0.01350]	0.01143	0.01135
64	2025	0.01258	0.01261	[0.01067,0.01434]	0.01209	0.01203
65	2026	0.01335	0.01336	[0.01126,0.01527]	0.01281	0.01279
66	2027	0.01409	0.01412	[0.01180,0.01622]	0.01351	0.01349
67	2028	0.01512	0.01515	[0.01267,0.01742]	0.01449	0.01448
68	2029	0.01630	0.01635	[0.01367,0.01886]	0.01562	0.01562
69	2030	0.01722	0.01727	[0.01430,0.01994]	0.01648	0.01641
70	2031	0.01873	0.01878	[0.01551,0.02170]	0.01791	0.01784
71	2032	0.01996	0.02001	[0.01645,0.02322]	0.01907	0.01889
72	2033	0.02165	0.02170	[0.01786,0.02520]	0.02069	0.02045
73	2034	0.02313	0.02317	[0.01901,0.02705]	0.02207	0.02176
74	2035	0.02471	0.02477	[0.02012,0.02898]	0.02357	0.02327
75	2036	0.02682	0.02689	[0.02176,0.03154]	0.02557	0.02521
76	2037	0.02896	0.02904	[0.02343,0.03411]	0.02760	0.02725
77	2038	0.03191	0.03200	[0.02575,0.03761]	0.03040	0.03010
78	2039	0.03496	0.03506	[0.02813,0.04132]	0.03332	0.03332
79	2040	0.03800	0.03809	[0.03045,0.04513]	0.03619	0.03611
80	2041	0.04226	0.04234	[0.03384,0.04999]	0.04025	0.04008
81	2042	0.04451	0.04460	[0.03541,0.05309]	0.04231	0.04213
82	2043	0.04906	0.04918	[0.03890,0.05858]	0.04662	0.04638
83	2044	0.05567	0.05576	[0.04448,0.06635]	0.05300	0.05237
84	2045	0.06068	0.06077	[0.04833,0.07237]	0.05774	0.05666
85	2046	0.07018	0.07034	[0.05644,0.08309]	0.06690	0.06625
86	2047	0.08107	0.08124	[0.06556,0.09554]	0.07742	0.07646
87	2048	0.09089	0.09111	[0.07416,0.10687]	0.08695	0.08524
88	2049	0.10455	0.10477	[0.08607,0.12207]	0.10019	0.09582
89	2050	0.11555	0.11581	[0.09484,0.13506]	0.11081	0.10627

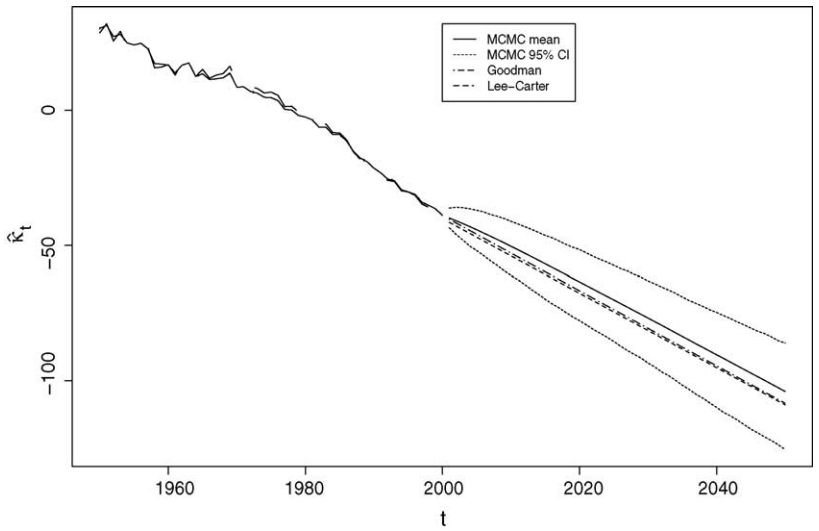


Fig. 7. Projection of the time factor κ_t to 2050, with MCMC confidence limits.

model used). Nevertheless, the Bayesian approach produces projected κ_t's above those furnished by the frequentist models, which in turn affects the mortality forecasts.

We can see in Table 3 the projected mortality rates for the generation aged 39 in year 2000. The MCMC credible intervals are somewhat large, and always include both Poisson Goodman and Lee–Carter projected values. The Bayesian point projections are nevertheless more pessimistic than those based on the Lee–Carter model: the mortality rates applying in the future to the generation aged 39 in year 2000 are always higher in the Bayesian approach. Note that the projected mortality rates derived from the Bayesian Poisson model also always exceed their Poisson Goodman counterparts. When applied to the pricing of life annuities, the projected lifetable obtained from the Poisson Bayesian model produces smaller premiums and reserves than the Poisson frequentist and Lee–Carter models.

Acknowledgements

The authors express their gratitude to the numerous comments made by the referee. These suggestions considerably improved the paper.

Claudia Czado greatly acknowledges the support through the SFB 386 "Statistical Analysis of discrete Structures" sponsored by the Deutsche Forschungsgemeinschaft.

Antoine Delwarde was supported by a research grant from the *Fonds de la Recherche pour l'Industrie et l'Agriculture* - FRIA. This financial support is gratefully acknowledged.

Michel Denuit thanks the support of Belgian Government under the "Projet d'Action de Recherches Concertées" 04/09-320.

References

Alho, J.M., 2000. Discussion of *Lee* (2000). N. Am. Actuarial J. 4, 91–93.
Brillinger, D.R., 1986. The natural variability of vital rates and associated statistics. Biometrics 42, 693–734.

Brooks, S.P., Roberts, G.O., 1998. Assessing convergence of Markov Chain Monte Carlo algorithms. Stat. Comput. 8, 319–335.
Brouhns, N., Denuit, M., Vermunt, J.K., 2002a. A Poisson log-bilinear approach to the construction of projected lifetables. Insur.: Math. Econ. 31, 373–393.
Brouhns, N., Denuit, M., Vermunt, J.K., 2002b. Measuring the longevity risk in mortality projections. Bull. Swiss Assoc. Actuaries, 105–130.
Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall/CRC.
Cowles, M.K., Carlin, B.P., 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. J. Am. Stat. Assoc. 91, 883–904.
Chib, S., 2001. Markov Chain Monte Carlo methods: computation and inference. Handbook of Econometrics, vol. 5. North Holland, pp. 381–394.
Dellaportas, P., Smith, A.F.M., Stavropoulos, P., 2001. Bayesian analysis of mortality data. J. Roy. Stat. Soc., Ser. A 164, 275–291.
Denuit, M., Lang, S., 2004. Nonlife ratemaking with Bayesian GAM's. Insur.: Math. Econ. 35, 627–647.
Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, New York.
Girosi, F., King, G., 2003. Demographic Forecasting. Monograph.
Goodman, L.A., 1979. Simple models for the analysis of association in cross-classifications having ordered categories. J. Am. Stat. Assoc. 74, 537–552.
Klugman, S., 1989. Bayesian modelling of mortality catastrophes. Insur.: Math. Econ. 8, 159–164.
Lee, R.D., 2000. The Lee–Carter method of forecasting mortality, with various extensions and applications. N. Am. Actuarial J. 4, 80–93.
Lee, R.D., Carter, L., 1992. Modelling and forecasting the time series of US mortality. J. Am. Stat. Assoc. 87, 659–671.
Renshaw, A.E., Haberman, S., 2003. Lee–Carter mortality forecasting with age specific enhancement. Insur.: Math. Econ. 33, 255–272.
Renshaw, A., Haberman, S., 2003. Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. Appl. Stat. 52, 119–137.
Sithole, T.Z., Haberman, S., Verrall, R.J., 2000. An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners' data. Insur.: Math. Econ. 27, 285–312.
Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. J. Roy. Stat. Soc., Ser. B 55, 3–23.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A., 2002. Bayesian measures of model complexity and fits (with discussion). J. Roy. Stat. Soc. B 64, 583–639.
Yashin, A.I., Iachine, I.A., Begun, A.S., 2000. Mortality modeling: a review. Math. Popul. Studies 8, 305–332.