

## A Bayesian Approach to Mixed Gamma Regression Models

Un enfoque bayesiano para modelos mixtos de regresión Gamma

MARTHA LUCÍA CORRALES-BOSSIO<sup>1,a</sup>, EDILBERTO CEPEDA-CUERVO<sup>2,b</sup>

<sup>1</sup>ESCUELA DE MATEMÁTICAS, UNIVERSIDAD SERGIO ARBOLEDA, BOGOTÁ, COLOMBIA

<sup>2</sup>DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ, COLOMBIA

---

### Abstract

Gamma regression models are a suitable choice to model continuous variables that take positive real values. This paper presents a gamma regression model with mixed effects from a Bayesian approach. We use the parametrization of the gamma distribution in terms of the mean and the shape parameter, both of which are modelled through regression structures that may involve fixed and random effects. A computational implementation via Gibbs sampling is provided and illustrative examples (simulated and real data) are presented.

**Key words:** Bayesian analysis; Gamma distribution; Gamma regression; Mixed models.

### Resumen

Los modelos de regresión gamma son una opción adecuada para modelar variables continuas que toman valores reales positivos. Este artículo presenta un modelo de regresión gamma con efectos mixtos desde un enfoque bayesiano. Utilizamos la parametrización de la distribución gamma en términos de la media y el parámetro de forma, los cuales se modelan a través de estructuras de regresión que pueden involucrar efectos fijos y aleatorios. Se proporciona una implementación computacional a través del muestreo de Gibbs y se presentan ejemplos ilustrativos (datos simulados y reales).

**Palabras clave:** Análisis bayesiano; Distribución Gamma; Regresión Gamma; Modelos mixtos.

---

<sup>a</sup>PhD. E-mail: [martha.corrales@usa.edu.co](mailto:martha.corrales@usa.edu.co)

<sup>b</sup>PhD. E-mail: [ecepedac@unal.edu.co](mailto:ecepedac@unal.edu.co)

## 1. Introduction

Mixed models are used in a wide variety of disciplines: physical, biological and social sciences, among others. These models have been particularly useful to represent clustered and, therefore, dependent data, arising for example when data are collected hierarchically, when observations are taken about related individuals, or when data are gathered over time regarding the same individuals (Ronquist & Huelsenbeck 2003, Vonesh 2006, Brown & Prescott 2014, Demidenko 2013). The study of hierarchical data has been largely framed for variables with normal distribution or at least symmetric distributions. However, models may not be appropriate when the response variable takes only positive values, such as, in the process of rate setting in the framework of heterogeneous insurance portfolios (vehicles, personal injury, etc.) (De Jong, Heller et al. 2008), where one possibility is to use a gamma regression model.

The substantial advantage of considering a gamma model is due to their flexibility compared to other models, such as exponential and Poisson, among others. Thus, gamma regression models allow for monotone models with no constant hazard in survival.

McCullagh & Nelder (1989) presented a gamma regression model where the coefficient of variation is assumed constant for all observations. Cepeda (2001) and Cepeda & Gamerman (2005) proposed an extension of the gamma regression models, assuming regression structures for both mean and dispersion parameters, and presented a Bayesian method to fit regression models in the two-parameter exponential family of distributions. Specifically for gamma observations, they modeled the mean and shape or variance parameters.

Following Cepeda (2001), our proposed model uses a parameterization of the gamma distribution in terms of the mean and the shape parameters. In this paper, we model the mean and the shape of a gamma distribution taking into account that, in this model, the variance is a function of the mean and shape parameters and that the mean and shape parameters are orthogonal in the Box-Cox sense (Cox & Reid 1987, Cepeda 2001). This property is not satisfied by the mean and variance.

The mean of the response variable is linked to a mixed-effects regression structure by the log and identity functions. An extended version of this model is also considered, assuming the shape parameter is not constant over the observations, but rather is related to a mixed-effects regression through the log link, similar to the mean.

This paper proposes a Bayesian method to fit gamma regression models, that include fixed and random effects in the mean and in the shape parameters. This paper is organized as follows: after an introduction, Section 2 present the Bayesian mixed gamma regression. In Section 3 a Bayesian method to fit the proposed models is presented. Section 4 show results of simulation studies. Section 5 include one application to real data set. Finally, in section 6 some conclusions are formulated. The outline of the OpenBUGS software (Thomas 2006) used, is presented in an appendice.

## 2. Bayesian mixed gamma regression models

Due of its flexibility and other characteristics, the gamma distribution is frequently used to model continuous data taking only positive values. The probability density function of a variable  $y$  that follows a gamma distribution, parameterized in terms of its mean  $\mu$  ( $\mu > 0$ ) and shape parameter  $\alpha$  ( $\alpha > 0$ ), is given by:

$$f(y|\mu, \alpha) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha y}{\mu}\right)^\alpha e^{-\alpha y/\mu} \left(\frac{1}{y}\right) I_{(0,\infty)}(y) \tag{1}$$

where  $\alpha > 0$  and  $\Gamma(\cdot)$  is the gamma function (Cepeda-Cuervo, 2001). In this case,  $E(y) = \mu$  and  $\text{Var}(y) = \frac{\mu^2}{\alpha}$ . If  $y$  has density function (1), we write  $y \sim G(\mu, \alpha)$ .

Now, let  $y_1, \dots, y_n$  be independent random variables such that  $y_i \sim G(\mu_i, \alpha)$ . The gamma regression model is defined assuming that

$$\eta_i = g(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is a vector of unknown regression parameters ( $p < n$ ),  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$  is a vector of covariates of the  $i$ -th observation and  $\eta_i$  is a linear predictor. Usually  $x_{i1} = 1$  for all  $i$ , so that the model has a mean intercept.

From generalized linear models, the canonical link function  $g(\cdot) : (0, \infty) \rightarrow \mathbb{R}$ , is the inverse function  $g(\mu) = 1/\mu$ . However, the usual link functions in the gamma regression models are  $g(\mu) = \log(\mu)$  and  $g(\mu) = \mu$ , due to the underlying constraints in the canonical link .

In the gamma regression model presented by Cepeda (2001), the shape parameter is not constant for every observation and it is modeled like the mean. Here, we assume  $n$  independent random variables,  $y_i \sim G(\mu_i, \alpha_i), i = 1, \dots, n$ , with mean and shape parameters given by

$$\eta_i = g(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta} \quad \tau_i = h(\alpha_i) = \mathbf{w}'_i \boldsymbol{\delta}$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  and  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)'$ ,  $p + k < n$ , are the sets of regression parameters,  $\mathbf{x}_i$  and  $\mathbf{w}_i$  are observed values of the covariates, and  $\eta_i$  and  $\tau_i$  are the linear predictors at the  $i$ -th observation.

Since the shape parameter is strictly positive, the log link function is a natural choice for  $h$ . It is then assumed that  $\log(\alpha_i) = \mathbf{w}'_i \boldsymbol{\delta}$  where  $\mathbf{w}_i$  is a vector of covariates and  $\boldsymbol{\delta}$  denotes a vector of unknown regression coefficients. Again, it is convenient to take the first component of the  $\mathbf{w}_i$  vector as 1 to allow for an intercept in the shape regression structure. There is no restriction on whether or not the  $\mathbf{w}_i$ 's contain the same predictor variables as the  $\mathbf{x}_i$ 's.

The gamma regression model described above does not involve random effects. To extend work in Bayesian generalized linear models and specifically in Bayesian gamma regression (Cepeda 2001, Cepeda & Gamerman 2004, Cepeda-Cuervo, Migon, Garrido & Achcar 2014), in this paper we propose two gamma mixed regression models, the first assuming that the shape parameter is constant for all observations and the second involving a mixed-effects model for the shape parameter.

Let  $\mathbf{y}_1, \dots, \mathbf{y}_m$  be independent continuous random vectors, where

$$\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$$

represents a vector of responses observed (with  $n_i$  measurements) for a sample unit  $i$  (in chronological order) and for which each of its components  $y_{ij}, j = 1, \dots, n_i$  takes positive real number values.

Also consider a regression model with the following structure:

$$G[E(\mathbf{y}_i|\mathbf{b}_i)] = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, \quad (2)$$

with  $i = 1, \dots, m$ , where  $G(\cdot)$  is a vector-function, linking the response vector of the conditional mean with the mixed linear model

$$\eta_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$$

where  $\mathbf{X}_i$  is the  $n_i \times p$  design matrix,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  are regression coefficients or fixed effects, and  $\mathbf{Z}_i$  is the  $n_i \times q$  design matrix associated with the vector of random effects  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})'$ .

For the identity link function, the  $j$ -th component of (2) is

$$\mu_{ij} = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i \quad (3)$$

where  $\mu_{ij} = E(y_{ij}|\mathbf{b}_i)$ ,  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})'$  and  $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ijq})'$ .

For the log link function, the  $j$ -th component of (2) is:

$$\log(\mu_{ij}) = \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i$$

where  $\mu_{ij} = E[y_{ij}|\mathbf{b}_i]$ ,  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})'$  and  $\mathbf{z}_{ij} = (z_{ij1}, \dots, z_{ijq})'$ , which is equivalent to

$$\mu_{ij} = \exp(\eta_{ij}) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) \quad (4)$$

In this paper, we first assume that conditional on  $\mathbf{b}_i$ ,  $\boldsymbol{\beta}$  and  $\alpha$ , that  $y_{ij}, i = 1, \dots, m; j = 1, \dots, n_i$ , are independent and have probability density function

$$y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \alpha \stackrel{ind}{\sim} G(\mu_{ij}, \alpha)$$

given by (1) with  $\mu$  replaced by  $\mu_{ij}$  (specified by (3) or (4)), depending on a set of explanatory variables through the selected link. In this formulation,  $\alpha$  represents a constant shape parameter.

In mixed models, the random effects,  $\mathbf{b}_1, \dots, \mathbf{b}_m$ , are typically assumed to be independent and normally distributed:

$$\mathbf{b}_i|\boldsymbol{\Sigma}_b \stackrel{ind}{\sim} N_q(0, \boldsymbol{\Sigma}_b), i = 1, \dots, m,$$

where  $\boldsymbol{\Sigma}_b$  is a  $q \times q$  positive-definite matrix. The normality assumption may be questionable in some practical situations when there are outliers, when the data

exhibit fat tails or the behavior of the data turns out to be asymmetric. An alternative is to consider other types of distributions such as the multivariate t-distribution with  $\nu_b > 0$  degrees of freedom, location vector  $\boldsymbol{\mu}_b \in \mathbb{R}^q$  and positive-definite dispersion matrix  $\boldsymbol{\Sigma}_b$  to model the random effects  $\mathbf{b}_i$ 's, i.e.,  $\mathbf{b}_i | \nu_b, \boldsymbol{\Sigma}_b \stackrel{ind}{\sim} t_q(\nu_b, \mathbf{0}, \boldsymbol{\Sigma}_b), i = 1, \dots, m$  (Figueroa-Zúñiga, Arellano-Valle & Ferrari 2013).

For a more general formulation of this model, we consider a different shape parameter,  $\alpha_{ij}$ , for each response  $y_{ij}$ . Let us now assume a mixed model for the logarithm of  $\alpha_{ij}$ . This is

$$\log(\alpha_{ij}) = \tau_{ij} = \mathbf{w}'_{ij} \boldsymbol{\delta} + \mathbf{h}'_{ij} \mathbf{d}_i \tag{5}$$

where  $\mathbf{w}_{ij} = (w_{ij1}, \dots, w_{ijp^*})'$  is the design vector corresponding to the vector  $p^* \times 1, \boldsymbol{\delta}$ , of fixed effects, and  $\mathbf{h}_{ij} = (h_{ij1}, \dots, h_{ijq^*})'$  is the design vector corresponding to the vector,  $\mathbf{d}_i$ , of random effects. It may be noted that the matrices  $\mathbf{W}_i = (w_{i1}, \dots, w_{in_i})$  and  $\mathbf{H}_i = (h_{i1}, \dots, h_{in_i})$  can contain the same covariate matrices  $\mathbf{X}_i = (x_{i1}, \dots, x_{in_i})$  and  $\mathbf{Z}_i = (z_{i1}, \dots, z_{in_i})$ , but this is not required.

Here it can be assumed that  $\mathbf{d}_i | \boldsymbol{\Sigma}_b \stackrel{ind}{\sim} N_q(\mathbf{0}, \boldsymbol{\Sigma}_b), i = 1, \dots, m$ , where  $\boldsymbol{\Sigma}_b$  is a positive-definite matrix. Alternatively, we can assume that  $\mathbf{d}_i | \boldsymbol{\Sigma}_d \stackrel{ind}{\sim} t_q(\nu_d, \mathbf{0}, \boldsymbol{\Sigma}_d), i = 1, \dots, m$ .

In order to apply Bayesian methods to fit the gamma Bayesian mixed model, we assume multivariate normal prior distributions for the fixed effects, that is

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

Vague priors are specified by taking large values for the prior variances. An alternative strategy is to consider a multivariate t-distribution, i.e.,

$$\boldsymbol{\beta} \sim t_p(\nu_\beta, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

and to specify an appropriate value for  $\nu_\beta$ , the degrees of freedom parameter. If the vector of random effects is assumed to follow a multivariate t-distribution, i.e.,  $\mathbf{b}_i | \nu_b, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b \sim t_q(\nu_b, \mathbf{0}, \boldsymbol{\Sigma}_b)$ , then the prior distribution for the degrees of freedom can be discrete as in Albert & Chib (1993) and Besag, Green, Higdon & Mengersen (1995), or continuous as in Geweke (1992). We have chosen the latter alternative. More specifically, we consider an exponential prior distribution with mean  $1/a$ , for the degrees of freedom, which we denote  $\varepsilon(a)$ . The prior distribution for the scale matrix of random effects  $\boldsymbol{\Sigma}_b$  is chosen, mainly for computational simplicity, to be an inverted Wishart distribution as in Fong, Rue & Wakefield (2010), i.e.,  $\boldsymbol{\Sigma}_b \sim IW_q(\boldsymbol{\Psi}, c)$ .

In summary, in this paper we study two gamma mixed regression models, where:

1. Model 1: The mean follows a mixed regression structure given by equation (2) modeled as in (3) or (4) (Model 1A and Model 1B, respectively) and both have a constant shape parameter  $\alpha$ .

2. Model 2: The mean follows a mixed regression structures given by (2) and the shape follows a regression structures given by (5).

### 3. Fitting Bayesian Mixed Gamma Regression Models

Let  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$  where  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})'$ , and  $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \dots, \boldsymbol{\eta}'_m)'$ , where  $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{in_i})'$ . By assumption, conditionally on  $\boldsymbol{\beta}, \boldsymbol{\Sigma}_b$  and  $\nu_b$ , the  $\boldsymbol{\eta}_i$ 's are independents and have density function  $f(\boldsymbol{\eta}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b) \propto f(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b)$ ,  $i = 1, \dots, m$ .

#### 3.1. Model 1

Assuming that the parameters  $\boldsymbol{\Sigma}_b, \nu_b, \alpha$  and  $\boldsymbol{\beta}$  are independents, the joint posterior distribution is given by:

$$f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \alpha, \boldsymbol{\eta} | \mathbf{y}) \propto \left[ \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \alpha) \right] \left[ \prod_{i=1}^m f(\eta_{ij} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b) \right] f(\boldsymbol{\Sigma}_b) f(\nu_b) f(\alpha) f(\boldsymbol{\beta})$$

Thus, samples of  $f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \alpha, \boldsymbol{\eta} | \mathbf{y})$  are obtained by an iterative process from the full conditional distributions:

$$f(\boldsymbol{\Sigma}_b | \nu_b, \boldsymbol{\beta}, \alpha, \boldsymbol{\eta}, \mathbf{y}), f(\nu_b | \boldsymbol{\Sigma}_b, \boldsymbol{\beta}, \alpha, \boldsymbol{\eta}, \mathbf{y}), f(\boldsymbol{\beta} | \boldsymbol{\Sigma}_b, \nu_b, \alpha, \boldsymbol{\eta}, \mathbf{y}) \\ f(\alpha | \nu_b, \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \boldsymbol{\eta}, \mathbf{y}), f(\boldsymbol{\eta}_i | \boldsymbol{\eta}_k, \boldsymbol{\Sigma}_b, \boldsymbol{\beta}, \alpha, \nu_b, \mathbf{y}), i, k = 1, \dots, m, i \neq k$$

The algorithm can be implemented using OpenBUGS and this software can also be used to obtain posterior parameter inferences.

#### 3.2. Model 2

Defining  $\boldsymbol{\tau} = (\boldsymbol{\tau}'_1, \dots, \boldsymbol{\tau}'_m)'$ , where  $\boldsymbol{\tau}_i = (\tau_{i1}, \dots, \tau_{in_i})'$ , and assuming (conditional independence of  $y_{ij} | \boldsymbol{\delta}, \boldsymbol{\Sigma}_d$  and  $\nu_d$ , the  $\boldsymbol{\tau}_i$ 's are independent and have density function  $f(\boldsymbol{\tau}_i | \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d) \propto f(\mathbf{d}_i | \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d)$ ,  $i = 1, \dots, m$ , and assuming independent prior distribution for  $\boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\delta}$  and  $\boldsymbol{\beta}$ , the posterior distribution of  $(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d)$  is given by:

$$f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau} | \mathbf{y}) \propto \left[ \prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \tau_{ij}) \right] \left[ \prod_{i=1}^m f(\boldsymbol{\eta}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b) \right] \\ \times \left[ \prod_{i=1}^m f(\boldsymbol{\tau}_i | \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d) \right] f(\boldsymbol{\Sigma}_b) f(\nu_b) f(\boldsymbol{\delta}) f(\boldsymbol{\Sigma}_d) f(\nu_d) f(\boldsymbol{\beta})$$

Thus, a Gibbs sampling algorithm also can be used to generate samples from:

$$f(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau} | \mathbf{y})$$

through iteratively sampling from the following full conditional distributions:

$$f(\boldsymbol{\beta} | \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y}), f(\boldsymbol{\Sigma}_b | \boldsymbol{\beta}, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y})$$

$$f(\nu_b | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y}), f(\boldsymbol{\delta} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\Sigma}_d, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y})$$

$$f(\boldsymbol{\Sigma}_d | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \nu_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y}), f(\nu_d | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \boldsymbol{\eta}, \boldsymbol{\tau}, \mathbf{y})$$

$$f(\boldsymbol{\eta}_i | \boldsymbol{\eta}_k, \boldsymbol{\tau}_k, \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \mathbf{y}_i), f(\boldsymbol{\tau}_i | \boldsymbol{\tau}_k, \boldsymbol{\eta}_k, \boldsymbol{\beta}, \boldsymbol{\Sigma}_b, \nu_b, \boldsymbol{\delta}, \boldsymbol{\Sigma}_d, \nu_d, \mathbf{y}_i)$$

for  $i, k = 1, \dots, m$ , and  $i \neq k$ . The algorithm for this model can also be implemented in OpenBUGS, and posterior inferences for the parameters can be made with this software.

## 4. Simulation studies

We conducted a simulation study to examine how similar the estimates of the parameters of the models are, compared with the true values of the parameters.

### 4.1. Model type 1

In this section we consider two classes of models:

**Model 1A.** In this section, we assume the model:

$$y_{ij} | \mathbf{b}_i, \alpha, \boldsymbol{\beta} \sim G(\mu_{ij}, \alpha), i = 1, \dots, n, \text{ and } j = 1, \dots, 5,$$

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})'$ ,

$$\eta_{ij} = \mu_{ij} = \beta_1 + \beta_2 x_{ij2} + \beta_3 x_{ij3} + b_{i1} + b_{i2} z_{ij2},$$

and  $\mathbf{b}_i | \nu_b, \boldsymbol{\Sigma}_b \sim t_2(\nu_b, \mathbf{0}, \boldsymbol{\Sigma}_b)$ , assuming as link function the inverse of the canonical link.

For  $n = 50, 100$  and  $150$ , values of the explanatory variables  $\mathbf{X}_2, \mathbf{X}_3$  and  $\mathbf{Z}_2$  were generated from uniform distributions  $U[0,30]$ ,  $U[0,15]$  and  $U[10,20]$  respectively, and we set  $\nu_b = 5, \alpha = 13, \boldsymbol{\beta} = (1.5, 2.0, 3.0)'$  and

$$\boldsymbol{\Sigma}_b = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}.$$

The following prior specifications were adopted:  $\nu_b \sim \varepsilon(0.1)$ ,  $\boldsymbol{\Sigma}_b \sim IW_2(\boldsymbol{\Psi}, c)$ , and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)' \sim t_3(\nu_\beta, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$  with  $c = 5$ , and  $\alpha \sim IG(\epsilon, \epsilon)$ , with  $\epsilon = 0.001$ ,

$$\boldsymbol{\Psi} = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix}, \quad \nu_\beta = 5, \quad \boldsymbol{\mu}_\beta = (0, 0, 0)', \quad \boldsymbol{\Sigma}_\beta = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{bmatrix}$$

The true values of the parameters and their estimations and standard deviations are presented in Table 1. In this model we consider 100,000 iterations, discarding the first 10,000 iterations as burn-in. The parameter estimates improve and standard deviations decline as sample size increases.

TABLE 1: True and estimated parameter values, and standard deviations for model 1A.

$n$		$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
	True	1.50	2.00	3.00	13.00	5.00	1.00	0.10	0.30
50	Estimated s.d.	1.23	2.10	3.13	17.77	4.92	0.93	0.12	0.27
		0.53	0.05	0.10	2.03	1.72	0.11	0.02	0.06
100	Estimated s.d.	1.39	2.07	3.15	15.25	5.19	1.03	0.11	0.29
		0.33	0.05	0.09	2.13	1.62	0.12	0.02	0.07
150	Estimated s.d.	1.53	2.05	3.11	15.37	4.95	1.03	0.10	0.31
		0.31	0.04	0.09	2.10	1.61	0.11	0.02	0.05

To monitor the chains' global convergence, we calculated the potential scale reduction factor (psrf) proposed by Brooks & Gelman (1998) (the psrf is an estimated factor by which the scale of the current distribution for the target distribution might be reduced if the simulations were continued for an infinite number of iterations). When the psrf is high (perhaps greater than 1.1 or 1.2), then we should run our chains longer to improve convergence to the stationary distribution. We obtained psrf = 0.995, 0.996 and 0.99 for  $n = 50, 100$  and  $150$ , respectively. The tests indicated that the Markov chain converged to its stationary distribution. Also, jumps were performed every 5 iterations to remove effects of autocorrelations in each chain.

The study of the convergence of the individual chains was performed using the CODA package (Plummer, Best, Cowles & Vines 2006), in the R software (R Core Team 2017). This package provides different diagnostic methods to check convergence: Gelman and Rubin's diagnostic (Geweke 1992), Geweke's diagnostic (Geweke 1992), Heidelberg and Welch's diagnostic (Heidelberger & Welch 1981) and Lewists diagnostic (Raffery & Lewis 1992).

Table 2 shows, for  $n = 50$ , the posterior correlations between the posterior parameters chains. Small correlations can be seen between them, with values between  $-0.11$  and  $0.15$ . Note that the correlations between  $\beta_i$  and  $\alpha$  are near zero. The correlations considering the other two sample sizes were similar to those shown in Table 2.

TABLE 2: Posterior correlations, model 1A.

	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
$\beta_1$	1.00							
$\beta_2$	-0.03	1.00						
$\beta_3$	-0.05	-0.07	1.00					
$\alpha$	-0.03	0.00	0.00	1.00				
$\nu_b$	-0.01	0.01	0.01	0.01	1.00			
$\Sigma_{b_{11}}$	0.06	-0.06	0.03	-0.07	-0.01	1.00		
$\Sigma_{b_{12}}$	-0.06	0.03	0.02	-0.01	-0.03	-0.05	1.00	
$\Sigma_{b_{13}}$	0.00	-0.01	0.00	-0.11	-0.08	0.02	0.15	1.00



**Model 1B.** We assume the model:

$$y_{ij}|\mathbf{b}_i, \alpha, \boldsymbol{\beta} \sim G(\mu_{ij}, \alpha), i = 1, \dots, n, \text{ and } j = 1, \dots, 5,$$

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})'$ ,

$$\eta_{ij} = \log(\mu_{ij}) = \beta_1 + \beta_2 x_{ij2} + \beta_3 x_{ij3} + b_{i1} + b_{i2} z_{ij2},$$

and  $\mathbf{b}_i|\nu_b, \boldsymbol{\Sigma}_b \sim t_2(\nu_b, 0, \boldsymbol{\Sigma}_b)$ .

For  $n = 50, 100$  and  $150$ , values of the explanatory variables  $\mathbf{X}_2, \mathbf{X}_3$  and  $\mathbf{Z}_2$  were generated, again, from uniform distributions  $U[0,30]$ ,  $U[0,15]$  and  $U[10,20]$ , respectively, and we set  $\nu_b = 5, \alpha = 13, \boldsymbol{\beta} = (1, 0.2, -0.03)'$ , and

$$\boldsymbol{\Sigma}_b = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 2 \end{bmatrix}.$$

Prior distributions for the parameters  $\nu_b, \boldsymbol{\Sigma}_b, \boldsymbol{\beta}$  and  $\alpha$  are the same as those proposed for model 1A.

The parameter estimations and standard deviations are given in Table 3. We consider 100,000 iterations, discarding the first 10,000 iterations as burn-in, like in the above model. Again we analyze the convergence of the chains with the above mentioned criteria. These diagnostic tests suggest good multivariate (psrf = 1.02) and individual behavior of the chains.

TABLE 3: True parameter values and standard deviations for model 1B.

$n$	True	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
50	Estimated	0.95	0.17	-0.03	13.00	5.00	0.80	0.40	2.00
	s.d.	0.37	0.09	0.01	2.03	0.67	0.11	0.10	0.76
100	Estimated	0.95	0.18	-0.03	17.75	5.19	1.03	0.41	1.89
	s.d.	0.33	0.08	0.01	2.01	0.62	0.12	0.10	0.67
150	Estimated	0.97	0.18	-0.03	15.37	4.95	1.03	0.41	1.91
	s.d.	0.31	0.09	0.01	2.10	0.61	0.11	0.09	0.65

Table 4 shows for  $n = 50$ , posterior correlations between the estimated parameters. As above, small correlations can be seen, with values between -0.13 and 0.16. Note, again, that the correlation between samples of  $\beta_2, \beta_3$  and  $\alpha$  is equal to zero. The correlations considering the other two sample sizes were similar to those shown in Table 4.

## 4.2. Model 2

In this section we assume the model

$$y_{ij}|\mathbf{b}_i, \boldsymbol{\delta}, \boldsymbol{\beta}, \mathbf{d}_i \sim G(\mu_{ij}, \alpha_{ij}), i = 1, \dots, n, \text{ and } j = 1, \dots, 5,$$

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})'$ ,  $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3)'$ ,  $\mathbf{d}_i = (d_{i1}, d_{i2})'$ , and mean and shape structures given by

TABLE 4: Posterior correlations, model 1B.

	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
$\beta_1$	1.00							
$\beta_2$	-0.02	1.00						
$\beta_3$	-0.01	-0.03	1.00					
$\alpha$	-0.06	0.00	0.00	1.00				
$\nu_b$	-0.02	0.02	0.03	0.03	1.00			
$\Sigma_{b_{11}}$	0.07	-0.06	0.03	-0.07	-0.01	1.00		
$\Sigma_{b_{12}}$	-0.09	0.06	0.04	-0.01	-0.04	-0.07	1.00	
$\Sigma_{b_{13}}$	0.03	-0.01	0.06	-0.13	-0.10	0.07	0.16	1.00

$$\eta_{ij} = \mu_{ij} = \beta_1 + \beta_2 x_{ij2} + \beta_3 x_{ij3} + b_{i1} + b_{i2} z_{ij2}$$

$$\tau_{ij} = \log(\alpha_{ij}) = \delta_1 + \delta_2 x_{ij2} + \delta_3 x_{ij3} + d_{i1} + d_{i2} z_{ij2},$$

and  $\mathbf{b}_i | \nu_b, \Sigma_{\mathbf{b}} \sim t_2(\nu_b, \mathbf{0}, \Sigma_{\mathbf{b}})$ .

We set  $\nu_b = 5$ ,  $\boldsymbol{\beta} = (0.5, 1.5, 2.5)'$ ,  $\boldsymbol{\delta} = (0.5, 0.2, 0.01)'$  and

$$\Sigma_{\mathbf{b}} = \Sigma_{\mathbf{d}} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 2 \end{bmatrix}.$$

For  $n = 50, 100$  and  $150$ , values of the explanatory variables  $\mathbf{X}_2$ ,  $\mathbf{X}_3$  and  $\mathbf{Z}_2$ , were generated from the uniform distributions defined in model 1A.

Prior distributions for the parameters  $\nu_b$ ,  $\Sigma_{\mathbf{b}}$  and  $\boldsymbol{\beta}$  are the same as those proposed for model 1A. Prior distribution for the parameter  $\boldsymbol{\delta}$  is the same as the parameter  $\boldsymbol{\beta}$ :  $t_3(\nu_{\beta}, \boldsymbol{\mu}_{\beta}, \Sigma_{\beta})$ . Also, for model specifications that include random effects for the shape parameter, we assume that the shape random effects  $\mathbf{d}_i$  have the same distribution as the location random effects  $\mathbf{b}_i$ :  $t_2(\nu_b, \mathbf{0}, \Sigma_{\mathbf{b}})$ .

In this model, we again considered 100,000 Monte Carlo iterations and the estimates were obtained using the samples obtained in the last 90,000 iterations. Testing the chains indicated very good convergence behavior. Table 6 shows, for  $n = 50$ , posterior correlations between posterior samples of the parameters. There are small correlations between them, with values between -0.15 and 0.17. Here, the correlations between posterior samples of  $\beta_2, \beta_3$ , and  $\delta_1, \delta_2, \delta_3$  are equal to zero. The correlations considering the other two sample sizes were similar to those shown in Table 6. The parameter estimates are given in Table 5.

We considered different values for  $\nu_b = 3, 7, 10$  for the models 1A, 1B and 2. The results obtained were good enough.

According to the results obtained in the above simulated models, the estimates obtained are reliable.

TABLE 5: True and estimated parameter values, and standard deviations for model 2.

$n$		$\beta_1$	$\beta_2$	$\beta_3$	$\delta_1$	$\delta_2$	$\delta_3$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
	True	0.50	1.50	2.50	0.50	0.20	0.01	5.00	0.80	0.40	2.00
50	Estimated	0.65	1.27	2.36	0.41	0.25	0.01	6.32	1.31	0.73	2.33
	s.d.	0.25	0.41	0.57	0.17	0.15	0.00	1.33	0.71	0.41	0.77
100	Estimated	0.61	1.29	2.41	0.43	0.21	0.01	6.09	0.93	0.68	2.11
	s.d.	0.15	0.35	0.52	0.13	0.11	0.00	1.22	0.53	0.39	0.75
150	Estimated	0.57	1.37	2.43	0.47	0.21	0.01	6.01	0.73	0.61	2.07
	s.d.	0.13	0.24	0.43	0.13	0.10	0.00	1.19	0.39	0.37	0.75

TABLE 6: Posterior correlations, model 2.

	$\beta_1$	$\beta_2$	$\beta_3$	$\delta_1$	$\delta_2$	$\delta_3$	$\nu_b$	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{13}}$
$\beta_1$	1.00									
$\beta_2$	-0.01	1.00								
$\beta_3$	-0.09	-0.03	1.00							
$\delta_1$	-0.04	0.00	0.00	1.00						
$\delta_2$	-0.06	0.00	0.00	0.00	1.00					
$\delta_3$	-0.03	0.00	0.00	0.00	0.00	1.00				
$\nu_b$	-0.01	-0.02	0.03	0.03	0.00	0.00	1.00			
$\Sigma_{b_{11}}$	-0.07	-0.09	0.08	0.03	0.02	0.01	0.01	1.00		
$\Sigma_{b_{12}}$	-0.03	0.08	0.05	-0.03	-0.03	-0.07	0.00	0.00	1.00	
$\Sigma_{b_{13}}$	-0.08	0.09	0.06	-0.15	-0.10	0.07	0.17	0.04	0.05	1.00

## 5. Application

Life insurance is a form of personal insurance that covers the risk of death of the insured, the occurrence of a serious illness or an unforeseen event that causes total and permanent disability of the insured.

The insurance data correspond to life insurance premiums (The values are premiums corresponding to contracts perfected or extended in the year, whose receipts have been issued in the corresponding period), in millions of Colombian pesos, of 19 insurance companies which operate the branches of life and person according to the Superintendence of Colombia from 2004 to 2015 (Source: Federation of Insurers of Colombia: FASECOLDA).

The data set is illustrated in Figure 1. This figure shows increasing behavior of the mean over time and that, conditional on the time, the data set presents a skewness like in a gamma distribution. Thus, giving the positivity of the data, it is assumed that the average life insurance premiums data can be modeled using a gamma distribution, with an appropriate link function.

### 5.1. Models with Constant Shape Parameter

Here, a mixed gamma regression model with a constant shape parameter is considered (Model 1), assuming location regression structure given by:  $\mu_{ij} = \beta_1 + \beta_2 t + b_{i1} + b_{i2} t$  (Model 1A and Model 1B) or  $\mu_{ij} = \beta_1 + \beta_2 t + \beta_3 t^2 + b_{i1} +$

$b_{i2}t + b_{i3}t^2$  (Model 1C and Model 1D)  $t = 1, 2, \dots, 12, i = 1, 2, \dots, 19$ . For the shape parameter  $\alpha$ , an inverse gamma distribution  $\alpha \sim IG(\epsilon, \epsilon)$ , with  $\epsilon = 0.001$  is assumed.

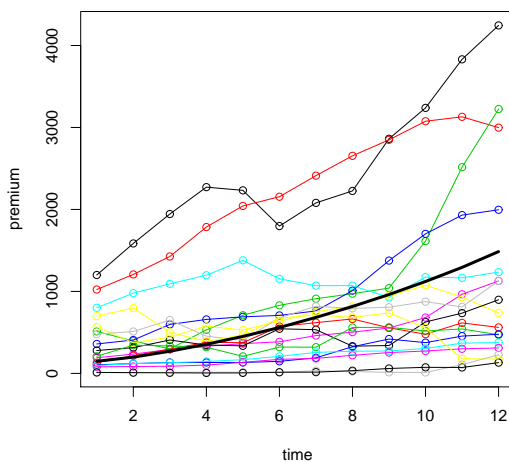


FIGURE 1: Premiums data. The darker line represents the fitted posterior mean. Model 1C

• **Model 1A:**

In this model, the following prior distributions were considered:  $\beta = (\beta_1, \beta_2)' \sim N_2(\mu_\beta, \Sigma_\beta)$  and  $\mathbf{b}_i = (b_{i1}, b_{i2})' \sim N_2(\mu_b, \Sigma_b)$ , with

$$\Sigma_\beta = \Sigma_b = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \mu_b = \mu_\beta = (0, 0)'.$$

• **Model 1B:**

In this model, the following prior specification was assumed:  $\beta = (\beta_1, \beta_2)' \sim t_2(\nu_\beta, \mu_\beta, \Sigma_\beta)$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})' \sim t_2(\nu_b, \mu_b, \Sigma_b)$ ,  $\nu_b \sim \varepsilon(a_0)$  and  $\Sigma_b \sim IW(\Psi, c)$ , with  $a_0 = 0.1, c = 5$ ,

$$\Sigma_\beta = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix}$$

and

$$\mu_b = \mu_\beta = (0, 0)'.$$

• **Model 1C:**

In this model, the following prior distributions were assumed:  $\beta = (\beta_1, \beta_2, \beta_3)' \sim N_3(\mu_\beta, \Sigma_\beta)$  and  $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})' \sim N_3(\mu_b, \Sigma_b)$ , with

$$\Sigma_{\beta} = \Sigma_{\mathbf{b}} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\mu}_{\mathbf{b}} = \boldsymbol{\mu}_{\beta} = (0, 0, 0)'$$

• **Model 1D:**

In this model, the following prior distributions were assumed:  $\beta = (\beta_1, \beta_2, \beta_3)' \sim t_3(\nu_{\beta}, \boldsymbol{\mu}_{\beta}, \Sigma_{\beta})$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})' \sim t_3(\nu_b, \boldsymbol{\mu}_b, \Sigma_b)$ ,  $\nu_b \sim \varepsilon(a_0)$  and  $\Sigma_b \sim IW(\Psi, c)$ , with  $a_0 = 0.1, c = 5$ ,

$$\Sigma_{\beta} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{bmatrix},$$

and

$$\boldsymbol{\mu}_b = \boldsymbol{\mu}_{\beta} = (0, 0, 0)'$$

TABLE 7: Estimated posterior medians and means, 95% credibility intervals (CI) for the parameters in Model 1C.

Parameter	Posterior inference			
	Mean	MC error	Median	95%CI
$\beta_1$	85.33	1.526	86.52	(62.7,110.0)
$\beta_2$	50.70	1.831	47.06	(26.5,75.14)
$\beta_3$	5.33	0.016	0.32	(0.38,12.19)
$\alpha$	1.85	0.021	1.83	(1.35,2.49)

These models were fitted using the OpenBugs program given in the Appendix. The correspondign DIC values for the fitted models are given by: for Model 1A, DIC=2305; for Model 1B, DIC=2376; for Model 1C, DIC=2286 and, for Model 1D, DIC=2293. Table 7 shows the posterior parameter estimates of Model 1C, which provide the least DIC value.

For the posterior samples of Model 1C (Figure 2), the multivariate version of Gelman and Rubin’s convergence diagnostic (Brooks and Gelman, 1998) indicates chain convergence (psrf = 0.98 < 1.2).

## 5.2. Models with Non-Constant Shape Parameter

Here a mixed gamma regression model is considered, assuming that  $\alpha$  is not constant through the observations and that its behavior can be explained by a time quadratic or a lineal relation. In this section, we consider four different prior specifications for the parameters of models: 2A, 2B, 2C and 2D: Cepeda

• **Model 2A**

In this model, we assume that mean and shape have regression structures are given by

$$\mu_{ij} = \beta_1 + \beta_2 t + b_{i1} + b_{i2} t \quad (6)$$

$$\log(\alpha_{ij}) = \delta_1 + \delta_2 t + d_{i1} + d_{i2} t \quad (7)$$

with the following prior specifications:  $\boldsymbol{\beta} = (\beta_1, \beta_2)' \sim N_2(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})' \sim N_2(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ , with

$$\boldsymbol{\Sigma}_\beta = \boldsymbol{\Sigma}_b = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \boldsymbol{\mu}_b = \boldsymbol{\mu}_\beta = (0, 0)'$$

and  $\boldsymbol{\delta} = (\delta_1, \delta_2)' \sim N_2(\boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta)$ ,  $\mathbf{d}_i = (d_{i1}, d_{i2})' \sim N_2(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$ , with

$$\boldsymbol{\Sigma}_\delta = \boldsymbol{\Sigma}_d = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \boldsymbol{\mu}_\delta = (0, 0)'$$

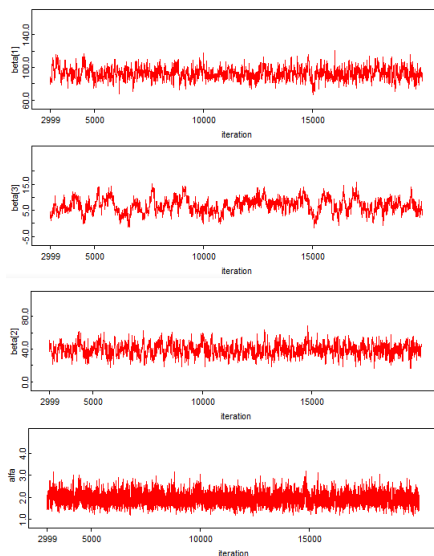


FIGURE 2: Behavior of the sample path. Model 1C.

#### • Model 2B

In this model we assume the mean and shape regression structures given by (6) and (7), and the following prior distributions:  $\boldsymbol{\beta} = (\beta_1, \beta_2)' \sim t_2(\nu_\beta, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2})' \sim t_2(\nu_b, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ ,  $\nu_b \sim \varepsilon(a_0)$ ,  $\boldsymbol{\Sigma}_b \sim IW_2(\boldsymbol{\Psi}, c)$ , with  $a_0 = 0.1$ ,  $c = 5$ ,  $\nu_\beta = 10$ , and

$$\boldsymbol{\Sigma}_\beta = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix}$$

and  $\boldsymbol{\mu}_b = \boldsymbol{\mu}_\beta = (0, 0)'$ ,  $\boldsymbol{\delta} = (\delta_1, \delta_2)' \sim t_2(\nu_\delta, \boldsymbol{\mu}_\delta, \boldsymbol{\Sigma}_\delta)$ ,  $\mathbf{d}_i = (d_{i1}, d_{i2})' \sim t_2(\nu_d, \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$ ,  $\nu_d \sim \varepsilon(a_0)$ ,  $\boldsymbol{\Sigma}_d \sim IW_2(\boldsymbol{\Psi}_1, c)$ , with  $a_0 = 0.1$ ,  $c = 5$ ,  $\nu_\delta =$

10, and

$$\Sigma_{\delta} = \begin{bmatrix} 1000 & 0 \\ 0 & 1000 \end{bmatrix}, \quad \Psi_{\mathbf{1}} = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix} \quad \mu_{\mathbf{d}} = \mu_{\delta} = (0, 0)'$$

• **Model 2C**

In this model, we assume the following regression structures:

$$\mu_{ij} = \beta_1 + \beta_2 t + \beta_3 t^2 + b_{i1} + b_{i2} t + b_{i3} t^2 \tag{8}$$

$$\log(\alpha_{ij}) = \delta_1 + \delta_2 t + \delta_3 t^2 + d_{i1} + d_{i2} t + d_{i3} t^2 \tag{9}$$

and the following prior distributions:  $\beta = (\beta_1, \beta_2, \beta_3)' \sim N_3(\mu_{\beta}, \Sigma_{\beta})$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})' \sim N_3(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$ , with

$$\Sigma_{\beta} = \Sigma_{\mathbf{b}} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \mu_{\mathbf{b}} = \mu_{\beta} = (0, 0, 0)'$$

and  $\delta = (\delta_1, \delta_2, \delta_3)' \sim N_3(\mu_{\delta}, \Sigma_{\delta})$ ,  $\mathbf{d}_i = (d_{i1}, d_{i2}, d_{i3})' \sim N_3(\mu_{\mathbf{d}}, \Sigma_{\mathbf{d}})$ , with

$$\Sigma_{\delta} = \Sigma_{\mathbf{d}} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \mu_{\delta} = (0, 0, 0)'$$

• **Model 2D**

In this model we assume the mean and shape regression structures given by (8) and (9), and the following prior distributions:  $\beta = (\beta_1, \beta_2, \beta_3)' \sim t_3(\nu_{\beta}, \mu_{\beta}, \Sigma_{\beta})$ ,  $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3})' \sim t_3(\nu_{\mathbf{b}}, \mu_{\mathbf{b}}, \Sigma_{\mathbf{b}})$ ,  $\nu_{\mathbf{b}} \sim \varepsilon(a_0)$ ,  $\Sigma_{\mathbf{b}} \sim IW_3(\Psi, c)$ , with  $a_0 = 0.1, c = 5, \nu_{\beta} = 10$ , and

$$\Sigma_{\beta} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{bmatrix}$$

and  $\mu_{\mathbf{b}} = \mu_{\beta} = (0, 0, 0)'$ ,  $\delta = (\delta_1, \delta_2, \delta_3)' \sim t_3(\nu_{\delta}, \mu_{\delta}, \Sigma_{\delta})$ ,  $\mathbf{d}_i = (d_{i1}, d_{i2}, d_{i3})' \sim t_3(\nu_{\mathbf{d}}, \mu_{\mathbf{d}}, \Sigma_{\mathbf{d}})$ ,  $\nu_{\mathbf{d}} \sim \varepsilon(a_0)$ ,  $\Sigma_{\mathbf{d}} \sim IW_3(\Psi_{\mathbf{1}}, c)$ , with  $a_0 = 0.1, c = 5, \nu_{\delta} = 10$ , and

$$\Sigma_{\delta} = \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}, \quad \Psi_{\mathbf{1}} = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 20 \end{bmatrix} \quad \mu_{\mathbf{d}} = \mu_{\delta} = (0, 0, 0)'$$

These models were fitted using the OpenBugs program given in the Appendix. The correspondign DIC values for the fitted models are given by: for Model 2A, DIC = 3744; for Model 2B, DIC = 3520; for Model 2C, DIC = 3815 and, for

Model 2D,  $DIC = 3855$ . Table 8 gives the posterior estimates of the parameters associated with model 2B, which provide the least DIC value.

We considered 20000 Monte Carlo iterations (to secure convergence) and our results were obtained with the posterior samples obtained from last 17000 iterations. Additionally, we performed the diagnostic tests reported for the simulated data, all of which suggested suitable behavior of the chains. For Model 2B, similar diagnostic evidences were obtained ( $psrf = 1.07$ ).

TABLE 8: Estimated posterior medians and means, 95% credibility intervals (CI) for the parameters in model 2B.

Parameter	Posterior inference			
	Mean	MC error	Median	95%CI
$\beta_1$	228.3	0.104	228.5	(203.04,278.61)
$\beta_2$	8.60	0.152	5.7	(7.73,15.22)
$\beta_3$	1.5	0.189	1.4	(0.92,2.04)
$\beta_4$	0.5	0.137	0.4	(0.21,0.93)

Figure 1 shows the posterior mean estimation  $\hat{\mu}$  through the time and the estimated trajectory of each of the premiums. The results suggest that the premiums average tends to increase over time, in non-linear form, as obtained in the model through a quadratic polynomial time. The Figure 1 shows that the variances are not homogeneous and tend to increase over time and that this variance can be modeled by a quadratic or cubic time polynomial. Between the constant shape model (Model 1C) and the variable shape model (Model 2B), the DIC values suggest that the first model is the best.

## 6. Conclusions and Extensions

The implementation of the mixed gamma regression model for random effects that are normally distributed and non-normally distributed is very challenging. In this sense, this approach is more flexible than others, because one can easily implement it when the distribution of the random effects follows the Student-t, skew normal or another distribution, by using simple and accessible software such as OpenBUGS. Another advantage of this approach is the easy implementation for the imputation of missing values, a common situation in longitudinal data for which a classic approach is much more complicated.

In the application we choose the identity link for the mean parameter. We did not have problem with this selection. However, other options, like the log link, are possible.

We selected the exponential distribution for the prior of the degrees of freedom of the multivariate t distribution, but other distributions may be used. A posterior study using different distributions can be implemented, such as that proposed by Fonseca, Ferreira & Migon (2008), and a comparison can be made. In general, a study of sensitivity to the choice of priors can be done, selecting different types of distributions, such as hierarchical distributions, among others.



It also possible to perform the study by modeling the mean and variance, instead of modeling the mean and shape, and comparing the obtained results.

[Received: December 2017 — Accepted: November 2018]

## References

- Albert, J. H. & Chib, S. (1993), ‘Bayesian analysis of binary and polychotomous response data’, *Journal of the American statistical Association* **88**(422), 669–679.
- Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995), ‘Bayesian computation and stochastic systems’, *Statistical science* pp. 3–41.
- Brooks, S. P. & Gelman, A. (1998), ‘General methods for monitoring convergence of iterative simulations’, *Journal of computational and graphical statistics* **7**(4), 434–455.
- Brown, H. & Prescott, R. (2014), *Applied mixed models in medicine*, John Wiley & Sons.
- Cepeda-Cuervo, E., Migon, H. S., Garrido, L. & Achcar, J. A. (2014), ‘Generalized linear models with random effects in the two-parameter exponential family’, *Journal of Statistical Computation and Simulation* **84**(3), 513–525.
- Cepeda, E. (2001), Modelagem da variabilidade em modelos lineares generalizados, Ph.D. tesis, Universidade Federal do Rio do Janeiro.
- Cepeda, E. C. & Gamerman, D. (2004), ‘Bayesian modeling of joint regressions for the mean and covariance matrix’, *Biometrical journal* **46**(4), 430–440.
- Cepeda, E. & Gamerman, D. (2005), ‘Bayesian methodology for modeling parameters in the two parameter exponential family’, *Revista Estadística* **57**(168-169), 93–105.
- Cox, D. R. & Reid, N. (1987), ‘Parameter orthogonality and approximate conditional inference’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–39.
- De Jong, P., Heller, G. Z. et al. (2008), *Generalized linear models for insurance data*, Vol. 10, Cambridge University Press Cambridge.
- Demidenko, E. (2013), *Mixed models: theory and applications with R*, John Wiley & Sons.
- Figuroa-Zúñiga, J. I., Arellano-Valle, R. B. & Ferrari, S. L. (2013), ‘Mixed beta regression: A bayesian perspective’, *Computational Statistics & Data Analysis* **61**, 137–147.

- Fong, Y., Rue, H. & Wakefield, J. (2010), ‘Bayesian inference for generalized linear mixed models’, *Biostatistics* **11**(3), 397–412.
- Fonseca, T. C., Ferreira, M. A. & Migon, H. S. (2008), ‘Objective Bayesian analysis for the Student-t regression model’, *Biometrika* **95**(2), 325–333.
- Geweke, J. (1992), ‘Evaluating the accuracy of sampling-based approaches to calculation of moments (with discussion)’, *Bayesian Statistics* **4**, 169–193.
- Heidelberger, P. & Welch, P. D. (1981), ‘A spectral method for confidence interval generation and run length control in simulations’, *Communications of the ACM* **24**(4), 233–245.
- McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, Chapman & Hall.
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2006), ‘CODA: convergence diagnosis and output analysis for MCMC’, *R news* **6**(1), 7–11.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Raffery, A. & Lewis, S. (1992), ‘One long run with diagnostics: Implementation strategies for markov chain monte carlo’, *Statist. Sci* **7**, 493–497.
- Ronquist, F. & Huelsenbeck, J. P. (2003), ‘Mrbayes 3: Bayesian phylogenetic inference under mixed models’, *Bioinformatics* **19**(12), 1572–1574.
- Thomas, A. (2006), ‘Making bugs open’, *R news* **6**, 12–17.
- Vonesh, E. F. (2006), ‘Mixed models: Theory and applications’.

## Appendix A. Appendix: BUGS Codes for the Mixed Gamma Regression

This appendix presents the various pieces of BUGS code used for fitting the mixed gamma regression in the simulated data example.

### Model 1A

```

model
{
  for( i in 1 : m ) {
    for( j in 1 : n ) {
      Y[ i , j ] ~ dgamma(a1[i,j] ,a2[i,j])
      a1[i,j] <- phi
      a2[i,j] <- phi/mu[ i , j ]
      mu[ i , j ] <- inprod(x[i, j, ], beta[ ]) + inprod(z[i, j, ], b[i,1,])
    }
    b[i,1,1:q] ~ dmt(cerovec [ ] ,psi[ ],g11)
  }
}

```

```

}
g11~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ ],g12)
V1[1:p,1:p] <- inverse(V[ ])
psi[1:q,1:q] ~ dwish(R0[ ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
phiinv ~ dgamma(a00,a00)
phi<-1/phiinv
}

```

## Model 1B

```

model
{
for( i in 1 : m ) {
for( j in 1 : n ) {
Y[i , j] ~ dgamma(a1[i,j] ,a2[i,j])
a1[i,j] <- phi
a2[i,j] <- phi/mu[i , j]
log(mu[i , j]) <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
}
b[i,1,1:q] ~ dmt(cerovec [ ],psi[ ],g11)
}
g11~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ ],g12)
V1[1:p,1:p] <- inverse(V[ ])
psi[1:q,1:q] ~ dwish(R0[ ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
phiinv ~ dgamma(a00,a00)
phi<-1/phiinv
}

```

## Model 2

```

model
{
for( i in 1 : m ) {
for( j in 1 : n ) {
Y[i , j] ~ dgamma(a1[i,j] ,a2[i,j])
a1[i,j] <- phi[i,j]
a2[i,j] <- phi[i,j]/mu[i , j]
log(phi[i,j])<-inprod(x[i, j, ], delta[ ])+inprod(z[i, j, ], gama[i,1,])
mu[i , j] <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
}
b[i,1,1:q] ~ dmt(cerovec [ ],psi[ ],g11)
gama[i,1,1:q] ~ dmt(cerovec [ ],psi[ ],g11)
}
g11~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ ],g12)
delta[1:p] ~ dmt(alpha[ ], V1[ ],g12)
V1[1:p,1:p] <- inverse(V[ ])
psi[1:q,1:q] ~ dwish(R0[ ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
}

```