Generalized Linear Models with Random Effects in the Two-Parameter Exponential Family

Edilberto Cepeda-Cuervo

Universidad Nacional, Bogotá, Colombia

Helio S. Migon

Federal University of Rio de Janeiro, Brazil

Liliana Garrido

Universidad de los Andes, Bogotá, Colombia

Jorge A. Achcar

University of Sao Paulo, Brazil

Abstract

In this paper we develop a new class of double generalized linear models, introducing a random effect component in the link function describing the linear predictor related to the precision parameter. This is a useful procedure to take into account extra variability and also to make the model more robust. The Bayesian paradigm is adopted to make inference in this class of models. Samples of the joint posterior distribution are draw using standard MCMC procedures. Finally, we illustrate this algorithm by considering simulated and real data sets.

Key words: Double generalized linear models, Stochastic models, Bayesian methodology, Biparametric exponential family.

1. Introduction

Generalized linear models have been a standard class of models for data analysis since the appearance of the classical book of McGullah and Nelder (1989). Recently, many extensions have been proposed with special emphasis on heterogeneous model data. The most frequently used approaches to remedy this sort of drawback are through mixture models and by the inclusion of random effects.

A two-parameter exponential family of models that accommodates overdispersion is proposed by Dey, Gelfand and Peng (1997), heavily based on an orthogonal reparamerization. Special attention has been devoted to modeling the mean and variance simultaneously. During the nineties, the interest in Taguchi type methods led to many developments in joint modeling the mean and dispersion from designed experiments with the advantage of avoiding the data transformation usually employed to satisfy the assumptions underlying classical linear models (normality, constant variance and linearity) (Nelder and Lee, 1991). The process of quality improvement aims to minimize the product variation caused by different types of noise. Quality improvement must be implemented in the design stage via design of experiments to assess the sensitivity of different control factors that affect the variability and mean of the process. Nelder and Lee (1991) discuss how the main ideas of GLM can be extended to analyze Taguchi experiments.

Classical and Bayesian inference approaches have been developed for this class of models. A classical approach to model variance heterogeneity in the normal set-up was proposed by Aitkin (1987). He explores the orthogonality between the mean and the variance and proposes distinct link functions for the mean and variance, which are monotonic differentiable real functions. In this way, the parameters can be estimated from a classical alternating algorithm. A Bayesian methodology was proposed by Cepeda and Gamerman (2001, 2005) to fit this model, assuming normal prior distributions for both mean and variance parameters in the regression models. These results were explored by Cepeda and Achcar (2010), in the context of heteroscedastic nonlinear regression.

The goal of this paper is to introduce random effect components to take into account extra data variability. A general Bayesian methodology for inference in this class of models is presented, and a simulation study is discussed. Illustrative examples of jointly modeling the mean and the precision parameters for some members of the two-parameter exponential family are considered. In particular, we consider models based on normal, beta and gamma distributions.

The paper is organized as follows. In Section 2 the double stochastic generalized linear models with random effects are presented. Section 3 provides the main steps involved in the MCMC algorithm used to draw inferences from the proposed model. Section 4 shows results from a simulation experiment carried out to study the performance of the proposed algorithm to fit the models. This section includes examples of modeling: (i) mean and variance, which are orthogonal parameters, in normal regression models and (ii) non-orthogonal parameters of Gamma distributions. Finally, in Section 5, we present the results of the analysis of real data sets, where the variable of interest comes from normal, gamma and beta distributions.

2. Generalized Linear Model with Random Effects

Consider the two-parameter exponential family models, largely discussed by Gelfand and Dalal (1990) and Dey, Gelfand and Peng (1997), of the form:

$$p(y|\theta,\tau) = b(y)\exp\{\theta y + \tau T(y) - \rho(\theta,\tau)\}. \quad (1)$$

where, if y is continuous (resp. discrete), p is assumed to be a density (resp. probability mass function) with respect to Lebesgue measure (resp. with respect to counting measure). They showed that if (1) is integrable over $y \in Y$, and if $T(y)$ is convex, then for a common mean, $Var(y)$ increases in τ. The normal, gamma and beta distributions are examples of distributions in this family, with θ, τ and ρ as given in Table 1. In Table 1, μ and σ^2 denote respectively the mean and variance parameters, α is the precision parameter in the gamma distribution and ϕ is the precision parameter in the beta distribution. For the beta distribution, a new random variable, $\tilde{y} = \text{logit}(y)$, is assumed to include it in the biparametric exponential family.

A particular case of this family occurs when $\tau = 0$ and corresponds to the well-known one-parametric exponential family. Examples of the distributions belonging to this family of distributions are included in Table 1. As an example, for a Pois(λ) distribution, $\theta = \log(\lambda)$, $\tau = 0$ and $\rho(\theta,0) = \rho^{'}(\theta,0) = \rho^{''}(\theta,0) = \exp(\theta)$, where Pois($\lambda$) denotes a Poisson distribution with mean equal to λ.

The class of doubly stochastic generalized linear models with random effects (DGLM-RE) can now be presented. Let Y_i, $i = 1, 2, \ldots, n$, be independent random variables in the-two parameter exponential family. Then:

1. *The random components*: the components $\boldsymbol{Y}_i$ have, conditional to the random effect, a distribution belonging to the two-parameter exponential family, with $E(\boldsymbol{Y}_i) = \mu_i$ and $Var(\boldsymbol{Y}_i) = \text{diag}(\sigma_i^2)$.

Distribution	θ	τ	$T(y)$	$\rho(\theta,\tau)$	$\rho^{'}(\theta,\tau)$	$\rho^{''}(\theta,\tau)$
Normal	$\frac{\mu}{\sigma^2}$	$\frac{1}{2\sigma^2}$	$-y^2$	$\frac{1}{2}\theta^2\sigma^2+\frac{1}{2}\log(2\pi\sigma^2)$	$\theta\sigma^2$	σ^2
Gamma	$\frac{-\alpha}{\mu}$	α	$\log(y)$	$\log[\Gamma(\alpha)(-\frac{\alpha}{\mu})^{-\alpha}]$	μ	μ^2/α
Exponential	$-\lambda$	0	-	$\rho(\theta,0)=-log(-\theta)$	$\rho(\theta,0)^{'}=-\frac{1}{\theta}$	$\rho(\theta,0)^{''}=\frac{1}{\theta^2}$
Beta	$\mu\phi$	ϕ	$\log(1-y)$	$\log\left[\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))}\right]$		
Poisson	$\log(\lambda)$	0	-	$\exp(\theta)$	$\exp(\theta)$	$\exp(\theta)$
Binomial	$logit(\lambda)$	0	-	$n\log(1+\exp(\theta))$	$\frac{n\exp(\theta)}{1-\exp(\theta)}$	$\frac{n\exp(\theta)(1-2\exp(\theta))}{1-\exp(\theta)}$

Table 1: Distributions of the exponential family.

2. *The systematic component*: the linear predictor $\boldsymbol{\eta}=(\boldsymbol{\eta_1},\boldsymbol{\eta_2})$ given by $\boldsymbol{\eta_1}=\boldsymbol{x}_i^{'}\boldsymbol{\beta}$ and $\boldsymbol{\eta_2}=\boldsymbol{z'}_i\boldsymbol{\gamma}+\nu_i$, where $\boldsymbol{x}_i$ is the ith-vector of the mean explanatory variables, $\boldsymbol{z}_i$ is the i^{th}-vector of the variance explanatory variables, $\boldsymbol{\beta}=(\beta_1,\beta_2,...,\beta_p)^{'}$ is the vector of the mean parameters, $\boldsymbol{\gamma}=(\gamma_1,\gamma_2,...,\gamma_r)^{'}$ is the vector of the variance parameters, and $\nu_i\sim N(0,\sigma_\nu^2)$.

3. *The link between the random and systematic components*: $\mu_i=h^{-1}(\eta_{1i})$ and $\sigma_i^2=g^{-1}(\eta_{2i})$, where h and g are monotonic differentiable functions.

This very broad class of models includes many well-known models in the family of generalized linear models in the one-parameter and two-parameter exponential families. In this way, DGLM-RE can be viewed as an extension of Double Generalized Linear Models (Cepeda and Gamerman, 2001; Cepeda and Gamerman, 2005).

In the definition of the DGLM-RE models, joint modeling of the mean and variance parameters is assumed. However, it includes a joint modeling of the mean and some other parameter, such as the dispersion parameter in the cases of the gamma and beta distributions.

2.1. Special cases

1. **Heteroscedastic normal regression models with random effect**. When there is heterogeneity of the variance, the homoscedastic assumption of the random components of the classical linear models fails. In this case, it is convenient to consider an analysis with

explicit modeling of the variance including possible explanations through explanatory variables. Thus, as a first example of DGLM-RE, we consider the joint modeling of the mean and the heterogeneity of the variance in a normal regression model considering linear or nonlinear models for the mean and variance. In this case, the vector of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ is related to the mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$ by $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is the error vector assumed to have multivariate normal distribution $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathrm{diag}(\sigma_i^2))$. The mean vector $\boldsymbol{\mu}$ is modeled by $\mu_i = h(\mathbf{x_i}, \boldsymbol{\beta})$, $i = 1, 2, \ldots, n$, where h is a linear function of the parameters. The variance heterogeneity is modeled through a set of explanatory variables $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{ir})'$ by $\sigma_i^2 = g(\mathbf{z}_i, \boldsymbol{\gamma}, \nu_i)$, where $\nu_i \sim N(0, \sigma_\nu^2)$, $i = 1, 2, \ldots, n$, are independent and g is an appropriate real function (monotonic, differentiable, which takes into account the positiveness of the variance).

2. **Joint mean and dispersion gamma regression models with random effect**. Here we consider the case where the random variable of interest comes from a gamma distribution. If the random variable Y has a gamma distribution with mean $\mu = \alpha\lambda$, where $\alpha, \lambda > 0$, its density function is given by

$$p(y|\alpha,\ \lambda) = y^{-1} \exp[-\frac{1}{\lambda} y + \alpha \log(y) - \log(K(\alpha, \lambda))] \tag{2}$$

and belongs to the biparametric exponential family, given that $b(y) = 1/y$, $\theta = -\frac{1}{\lambda} = -\frac{\alpha}{\mu}$. The double generalized gamma random effect models are defined by

$$h(\mu) = \mathbf{x}_i' \boldsymbol{\beta} \quad \text{and} \quad g(\alpha_i) = \mathbf{z}_i' \boldsymbol{\gamma} + \nu_i. \tag{3}$$

Often the link functions $h(\mu_i) = \log(\mu_i)$ and $g(\alpha_i) = \log(\alpha_i)$ are chosen and assumed known.

3. **Joint mean and dispersion beta regression models with random effect**. As a third case, we consider the case where the random variable of interest comes from a beta distribution. Let Y be a beta distribution, $Y \sim B(\alpha, \lambda)$, with mean $\mu = \alpha/(\alpha + \lambda)$, where $\alpha, \lambda > 0$. Taking $\phi = \alpha + \lambda$, this density can be rewritten as

$$f(y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi - 1} (1 - y)^{(1-\mu)\phi - 1} I_{(0,1)}(y), \tag{4}$$

where $I_{(0,1)}(y)$ is the indicator function, which equals 1 if y belongs to the real open interval $(0,1)$ and 0 if not. This reparametrization that is presented in Ferrari and Cribari-Neto (2004) was introduced early in the literature, for example by Jorgensen (1997) and by Cepeda (2001). Joint beta regression models were proposed by Cepeda(2001) and by Cepeda and Gamerman(2005). In these works, joint modeling of the mean and dispersion parameters is proposed as $\text{logit}(\mu_i) = \mathbf{x_i}'\boldsymbol{\beta}$ and $\log(\phi_i) = \mathbf{z_i}\gamma$. This proposal was also studied by Smithson and Verkuilen (2006) and next studied by Simas et al. (2010). A nonlinear beta regression model was proposed by Cepeda and Achcar (2010). The double generalized beta random effect models are defined using the linear predictors given by

$$h(\mu) = \mathbf{x}_i'\boldsymbol{\beta} \quad \text{and} \quad g(\phi_i) = \mathbf{z}_i'\boldsymbol{\gamma} + \nu_i \tag{5}$$

Often the link functions $h(\mu_i) = \text{logit}(\mu_i)$ and $g(\phi_i) = \log(\phi_i)$ are chosen in each application. In this case, we can consider the joint modeling of the mean and variance parameters, assuming that $g(\sigma_i^2) = \mathbf{z}_i'\boldsymbol{\gamma} + \nu_i$, where g is an appropriate real function.

4. **Overdispersed regression models.** The class of DGLM-RE includes special cases of the well-known linear overdispersed models:

   a) **Poisson normal model**. This model assumes that $Y_i|\lambda_i \sim Pois(\lambda_i),\ i = 1, \cdots, n$, are conditionally independent, with mean model given by:

   $$\ln(\lambda_i) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \nu_i, \tag{6}$$

   where ν_i, $i = 1, 2, \ldots, n$, are random variables having normal distribution with mean zero and variance σ_ν^2, $\nu_i \sim N[0, \sigma_\nu^2]$ (See Hinde and Demetrio, 1998; and Quintero-Sarmiento, Cepeda-Cuervo and Nuñez-Anton, 2011).

   b) **Binomial normal model**. This model assumes that $Y_i|\pi_i \sim Bin(n, \pi_i)$, $i = 1, 2, ...n$, are conditionally independent random variables, with mean model given by

   $$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \nu_i, \tag{7}$$

   where ν_i, $i = 1, 2, \ldots, n$, are random variables having a normal distribution with mean zero and variance σ_ν^2, $\nu_i \sim N[0, \sigma_\nu^2]$ (Williams, 1982).

3. Bayesian inference

To fit the proposed models under the Bayesian paradigm, we use MCMC methods to simulate samples for the joint posterior distribution of interest. A first approach can be made using the WinBugs software with a slice generalization of the computer programs proposed by Cepeda and Achcar(2010). In a second approach, we apply a slight generalization of the Bayesian algorithm proposed for linear models by Cepeda and Gamerman(2001, 2005). For model comparison, we use the Deviance Information Criterion (DIC), where smaller values indicates better models (Spiegelhalter et al., 2002).

3.1. Proposed Bayesian methodology

The idea that we summarize in this section is a generalization of the algorithm presented in Cepeda and Gamerman (2001, 2005) as a Bayesian methodology to fit joint modeling of the mean and variance (or other parameter) in the biparametric exponential family. In this paper, as a special case, the mean and variance parameters are modeled as functions of the regression models $h(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ and $g(\sigma_i^2) = \boldsymbol{z}_i'\boldsymbol{\gamma} + \nu_i$, with h and g being appropriate real functions.

To estimate the parameters of this model using Bayesian methodology, independent normal prior distributions, $\boldsymbol{\beta} \sim N(\boldsymbol{b}, \boldsymbol{B})$ and $\boldsymbol{\gamma} \sim N(\breve{\boldsymbol{g}}, \boldsymbol{G})$, are assigned for the mean and variance parameters. Thus, from the Bayes theorem, the posterior distribution function is given by $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{Y}, \boldsymbol{X}) \sim L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{Y}, \boldsymbol{X})p(\boldsymbol{\beta}, \boldsymbol{\gamma})$, where $L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{Y}, \boldsymbol{X})$ denotes the likelihood function and $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ the joint prior density function. Given that, $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{Y}, \boldsymbol{X})$ is analytically intractable and, thus, it is not easy to generate samples from it, we propose sampling $(\boldsymbol{\beta}, \boldsymbol{\gamma}|\boldsymbol{Y}, \boldsymbol{X})$ in an iterated process, sampling $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from the conditional posterior distributions $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{Y}, \boldsymbol{X})$ and $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{X})$ respectively. But these distributions are also intractable and not easily generated, except in the case where h is the identity function and the observation comes from normal distributions (Cepeda and Gamerman, 2001). Therefore, we must build transition kernels which allow us to simulate samples that will be part of the posterior conditional distributions.

To get samples of the posterior conditional distribution $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{Y}, \boldsymbol{X})$, working observational variables are proposed as first-order Taylor approximation of $h(Y_i)$ around the current values of

$E(Y_i) = \mu_i^{(c)}$ given by the current values of the mean regression models $\boldsymbol{\beta}^{(c)}$. Thus, the working observations are given by

$$\tilde{Y}_i \;=\; h(\mu_i^{(c)}) + h^{'}(\mu_i^{(c)})(Y_i - \mu_i^{(c)}), \quad i = 1, 2, \ldots, n, \tag{8}$$

where $h(\mu_i^{(c)}) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}^{(c)}$.

This working variable has mean and variance given respectively by $E(\tilde{Y}_i) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}^{(c)}$ and $Var(\tilde{Y}_i) = [h^{'}(\mu_i^{(c)})]^2 Var(Y_i)$. Thus, the normal transition kernel is given by the posterior distribution resulting from the combination of the normal prior distribution $\boldsymbol{\beta} \sim N(\boldsymbol{b}, \boldsymbol{B})$ with the likelihood function, assuming that $\tilde{Y}_i$, $i = 1, 2, \ldots, n$, has a normal distribution. In this way, the kernel transition function is given by the normal distribution

$$q_1(\boldsymbol{\beta}|\boldsymbol{\beta}^{(c)}, \boldsymbol{\gamma}^{(c)}) = N(\boldsymbol{b}^*\ , \boldsymbol{B}^*), \tag{9}$$

where $\boldsymbol{b}^* = \boldsymbol{B}^*(\boldsymbol{B}^{-1}\boldsymbol{b} + \boldsymbol{X}^t\Sigma^{-1}\tilde{\boldsymbol{Y}})$, $\boldsymbol{B}^* = (\boldsymbol{B}^{-1} + \boldsymbol{X}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$ and $\boldsymbol{\Sigma} = \text{diag}(\sigma_i^2)$. Thus, values of $\boldsymbol{\beta}$ that will be part of the posterior distribution sample will be proposed from (9), in the Metropolis Hastings algorithm (Chib and Greenberg, 1995; Gamerman, 1997).

To get samples of the posterior conditional distribution $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{X})$, a working observational variable also is defined. For it, we first build a random variable T_i such that $E(\boldsymbol{T}_i) = \sigma_i^2$ and next we define the working variable using a first-order Taylor approximation of the real valued function $g(t_i)$. Thus, in this case, it is evident that $T_i = (Y_i - \mu_i)^2$ and the working observational variable is given by

$$\hat{Y}_i \;=\; g(\sigma_i^{2(c)}) + g^{'}(\sigma_i^{2(c)})(T_i - \sigma_i^{2(c)}), \quad i = 1, 2, \ldots, n, . \tag{10}$$

where $g(\sigma_i^{2(c)}) = \boldsymbol{z}_i^{'}\boldsymbol{\gamma}^{(c)}$.

This random variable has working observational mean and working observational variance given respectively by $E(\hat{Y}_i) = \boldsymbol{z}_i^{'}\boldsymbol{\gamma}^{(c)}$ and $Var(\hat{Y}_i) = [g^{'}(\sigma_i^{2(c)})]^2 Var(T_i)$. Thus, the normal transition kernel is given by the posterior distribution resulting from the combination of the normal prior distribution $\boldsymbol{\gamma} \sim N(\breve{\boldsymbol{g}}, \boldsymbol{G})$ with the likelihood function, assuming that $\hat{Y}_i$, $i = 1, 2, \ldots, n$, has a normal distribution. Thus, the kernel transition function is given by a normal distribution. That is,

$$q_2(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(c)}, \boldsymbol{\beta}^{(c)}) = N(\boldsymbol{g}^* \ , \boldsymbol{G}^*), \tag{11}$$

where $\boldsymbol{g}^* = \boldsymbol{G}^*(\boldsymbol{G}^{-1}\breve{\boldsymbol{g}} + \boldsymbol{Z}^t\Psi^{-1}\tilde{\boldsymbol{Y}})$, $\boldsymbol{G}^* = (\boldsymbol{G}^{-1} + \boldsymbol{Z}^t\Psi^{-1}\boldsymbol{Z})^{-1}$ and Ψ is a diagonal matrix with diagonal elements $\hat{\sigma}_i^2 = \text{Var}(Y_i)$, for $i = 1, 2, ..., n$. As a consequence, values of $\boldsymbol{\gamma}$ that will be part of the posterior sample of $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$ will be proposed from (11), in the Metropolis Hastings algorithm (Chib and Greenberg, 1995).

Table 2 presents working observational variables associated with the modeling of the parameters of distributions belonging to the biparametric exponential family. In the normal distribution, the working observations are the same as those resulting from the Fisher scoring algorithm. For gamma distributions, there are many options for parameter modeling, for example mean and variance or mean and shape parameters (Cepeda and Gamerman, 2001). For the beta distribution, working variables as included in Table 2 are proposed when the mean and "precision"parameter are being modeled, but also it is possible to build a general proposal to model mean and variance. In this way, the random variable T_i used to build the working variable $\hat{Y}_i$, can be defined easily. For example, in the case of a gamma distribution, given that $E(Y_i) = \alpha_i\lambda_i$ and $E(Y_i/\lambda_i) = \alpha_i$, the random variable T_i used to build the working observation is given by $T_i = Y_i/\lambda_i$. In the case of modeling the precision parameters of beta distributions, given that $E(Y_i) = \alpha_i/\phi_i$, the random variable T_i used to build the working observation is $T_i = \phi^2 Y_i/\alpha_i$ (Note that $E(T_i) = \phi_i$).

Values of $\varphi = \sigma_\nu^{-2}$ will be generated from the proposed density

$$\pi(\varphi) = \varphi^{\frac{n}{2}+p-1} \ \exp\{-(\lambda + \frac{1}{2}S_0^2)\varphi\} \tag{12}$$

where $S_o^2 = \sum_{i=1}^n \nu_i^2$. This posterior distribution results from assigning a prior distribution $p(\varphi) \propto \varphi^{p-1}\exp(-\lambda\varphi)$, given that $\nu_i \sim N(0, \sigma_\nu^2)$, where $\varphi = \sigma_\nu^{-2}$.

With the kernel transition functions (10), (11)and (12), an iterative algorithm can be proposed to simulate samples of the posterior distributions, in each one of the proposed models.

Distribution	Parameter model	T	Working Variable $\tilde{Y}$
Normal	$\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$	Y_i	Y_i
	$\log(\sigma_i^2) = \mathbf{z}_i'\boldsymbol{\gamma}$	$(Y_i - \mathbf{x}_i'\boldsymbol{\beta}^{(c)})^2$	$\mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{(Y_i - \mathbf{x}_i'\boldsymbol{\beta}^{(c)})^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)})} - 1$
Gamma	$\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$	Y_i	Y_i
	$\log(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$	Y_i	$\mathbf{x}_i'\boldsymbol{\beta}^{(c)} + \frac{Y_i}{\mu^{(c)}} - 1$
	$\log(\alpha_i) = \mathbf{z}_i'\boldsymbol{\gamma}$	Y_i/λ	$\mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{Y_i}{\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(c)})} - 1$
	$\log(\sigma_i^2) = \mathbf{z}_i'\boldsymbol{\gamma}$	λY_i	$\mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{Y_i - \mu^{(c)}}{\mu^{(c)}}$
Beta	$\text{logit}(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$	Y_i	$\mathbf{x}_i'\boldsymbol{\beta}^{(c)} + \frac{Y_i - \mu^{(c)}}{\mu^{(c)}(1-\mu^{(c)})}$
	$\log(\phi_i) = \mathbf{z}_i'\boldsymbol{\gamma}$	$\frac{\phi^2}{\alpha}Y_i$	$\mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{Y_i}{\mu_i^{(c)}}$
Poisson	$\log(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$	Y_i	$\mathbf{x}_i'\boldsymbol{\beta}^{(c)} + \frac{1}{\mu^{(c)}}(Y_i - \mu^{(c)})$
Binomial	$\text{logit}(\theta_i) = \mathbf{x}_i'\boldsymbol{\beta}$	Y/n	$\mathbf{x}_i\boldsymbol{\beta}^{(c)} + \frac{Y_i - n\theta^{(c)}}{n\theta^{(c)}(1-\theta^{(c)})}$

Table 2: Working Variables for each distribution.

3.2. MCMC algorithm

With these sampling proposals, the components $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and σ_ν^2 of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\nu^2)'$ are updated in these steps:

1. Begin the chain interactions counter $j = 1$ and give initial values $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$ and $\sigma_{\nu_0}^2$ for $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\nu^2)'$ and $\nu_0 = (\nu_{10}, \nu_{20}, \ldots, \nu_{n0})'$ for $\nu = (\nu_1, \nu_2, \ldots, \nu_n)$.
2. Move the vector $\boldsymbol{\beta}$ to a new value $\boldsymbol{\zeta}$ generated from the proposed density $q_1(\beta^{(j-1)}|.)$.
3. Calculate the acceptance probability of movement $\alpha(\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\zeta})$. If the movement is accepted, then $\boldsymbol{\beta}^{(j)} = \boldsymbol{\zeta}$. If it is not accepted, then $\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)}$.
4. Move the vector $\boldsymbol{\gamma}$ to a new value $\boldsymbol{\zeta}$, generated from the proposed density $q_2(\boldsymbol{\gamma}^{j-1}|.)$.
5. Calculate the acceptance probability of movement $\alpha(\boldsymbol{\gamma}^{(j-1)}, \boldsymbol{\zeta})$. If the movement is accepted, then $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\zeta}$. If it is not accepted, then $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{(j-1)}$.
6. Move φ to a new value ζ, generated from the proposed density

$$\pi(\varphi) = \varphi^{\frac{n}{2}+p-1} \ \exp\{-(\lambda + \frac{1}{2}S_0^2)\varphi\} \tag{13}$$

where $S_o^2 = \sum_{i=1}^n \nu_i^2$.

7. Calculate the acceptance probability of movement $\alpha(\varphi^{(j-1)}, \zeta)$. If the movement is accepted, then $\varphi^{(j)} = \zeta$. If it is not accepted, then $\varphi^{(j)} = \varphi^{(j-1)}$.

8. Update ν_i, i=1,2,...,n, using a random walk $\nu_i = \nu_i^{(c)} + \epsilon$, where $\epsilon \sim N(0, \tau)$, for a given τ.

9. Finally, update the counter from j to $j+1$ and return to 2 until convergence.

In this algorithm, given a current vector of regression parameters θ_V, $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$, a new vector of the parameters θ_N is proposed, with acceptance probability of the movement given by:

$$\begin{aligned}\alpha(\theta_V, \theta_N) \quad = \quad & \exp\{-\frac{1}{2^k}[(\theta_V - \theta_0)'(\theta_V - \theta_0) - \frac{1}{2^k}(\theta_N - \theta_0)'(\theta_N - \theta_0) \\ & + \log(L(\theta_V)) - \log(L(\theta_N)) \\ & + \frac{1}{2}(\theta_V - b_N^*)B_N^*(\theta_V - b_N^*) + \frac{1}{2}(\theta_N - b_V^*)B_V^*(\theta_N - b_V^*)\},\end{aligned}$$

where θ_0 is given by the mean of the normal prior distribution. The parameter values of b_V^*, b_N^*, B_V^* and B_N^* are given by the proposed transition kernel equations. In this notation, the subscript V is associated with the current values of the parameters and N with the proposed values, associated with the proposed movement of the chain. The capital letter $L(.)$ denotes the value of the likelihood function.

4. Simulated studies

In this section, we present some simulation data in order to evaluate the performance of the proposed algorithms in comparison with the WinBugs outputs. Both normal and gamma distributed data are considered with different degrees of random effects controlled by the corresponding variance. In the last simulated cases, the main objective is to show empirically how the presence of random effects compares to correlated observations. In all cases, independent normal prior distributions are assigned for the mean and variance (precision) regression parameters and a gamma prior distribution for the variance of the random effects.

4.1. Heteroscedastic normal regression models with random effects

In the first simulated study, two random explanatory variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ were considered. Values x_{1i} and x_{2i}, $i = 1, \ldots, 50$, of these variables were generated from uniform distributions: $U(0, 10)$ and $U(0, 20)$, respectively. Values of the variable of interest $\boldsymbol{Y}$ were generated from a normal distribution with mean $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ and variance $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_{2i} + \gamma_2 x_{3i} + \nu_i)$, where $\nu_i \sim N(0, \sigma_\nu^2)$. In this simulation, we considered three models, all with true regression value parameters given by $\beta_0 = -4.1$, $\beta_1 = 2.35$, $\beta_2 = 0.65$, $\gamma_0 = -6$, $\gamma_1 = 0.25$ and $\gamma_2 = -0.2$, but with different values of σ_ν^2: 0.2, 0.4, 0.6. In order to apply the Bayesian methodology developed in this paper, we assigned normal prior distributions for the location parameters, given by $\beta_k \sim N(0, 10^5)$, $\gamma_k \sim N(0, 10^2)$, $k = 0, 1, 2$, and a gamma prior distribution, for the variance of the random effect, $\sigma_\nu^2 \sim Gamma(0.01, 0.01)$.

For each of these simulations, many alternative chains were generated with different starting values. All of them converged to the same values after a small burn-in period of iterations and they exhibited the same qualitative behavior, providing a rough indication of stationarity. For each σ_ν^2, we fitted alternative models (with random effects, without random effects, without explanatory variables in the variance models), but in all cases, as expected, the model with the smallest DIC was that from where the data was generated.

The posterior parameter estimates obtained applying the proposed MCMC algorithm and using WinBugs software are given in Table 3. The first row of Table 3 shows the parameter estimates when the true value of σ_ν^2 is $\sigma_\nu^2 = 0.2$, obtained by applying the proposed MCMC algorithm, developed using MathLab software (PA). The second row shows the parameter estimates obtained using WinBugs, and so on. The results presented in Table 3 are based on one long chain of size 10000 and burn-in period of size 3000. For all the parameters, in all simulations, the parameter estimates are very close to the respective true values, all of them with small standard deviations.

Other simulations were performed with σ_ν^2 equal to 0.0, 0.2, 0.4, 0.8, 1.0, 2.0 or 3.0, with the same two explanatory variables in the mean and in the variance models. In these cases, the best model is either a model without X_1 or X_2, or a model with a random effect only.

Parameters		β_0	β_1	β_2	γ_0	γ_1	γ_2	σ^2_ν
$\sigma^2_\nu = 0.2$	Mean	-4.111	2.350	0.651	-5.201	0.235	-0.262	0.188
PA	s.d	0.005	0.001	0.000	0.400	0.055	0.031	0.178
$\sigma^2_\nu = 0.2$	Mean	-4.111	2.350	0.651	-5.274	0.242	-0.259	0.240
WB	s.d	0.005	0.001	0.000	0.411	0.055	0.031	0.301
$\sigma^2_\nu = 0.4$	Mean	-4.107	2.350	0.650	-5.511	0.159	-0.201	0.443
PA	s.d	0.006	0.001	0.001	0.553	0.058	0.035	0.338
$\sigma^2_\nu = 0.4$	Mean	-4.107	2.350	0.650	-5.867	0.165	-0.184	0.985
WB	s.d.	0.006	0.001	0.001	0.615	0.065	0.038	0.635
$\sigma^2_\nu = 0.6$	Mean	-4.095	2.350	0.649	-6.191	0.224	-0.177	0.702
PA	s.d	0.007	0.0019	0.000	0.483	0.061	0.032	0.466
$\sigma^2_\nu = 0.6$	Mean	-4.096	2.350	0.650	-6.544	0.245	-0.175	1.657
WB	s.d	0.006	0.001	0.000	0.601	0.078	0.038	0.870

Table 3: Posterior mean and standard deviation of the parameters of the normal random effect models.

4.2. Joint mean and dispersion gamma regression random effect models

In the second simulation study, the same two random explanatory variables X_1 and X_2 were considered, but at this time we only generated 40 observed values of the explanatory variables. The values y_i, $i = 1, 2, \ldots, 40$, of the variable of interest Y were generated from a gamma distribution with mean $\mu_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$ and variance $\sigma^2_i = \exp(\gamma_0 + \gamma_1 x_{2i} + \gamma_2 x_{3i} + \nu_i)$, where $\nu_i \sim N(0, \sigma^2_\nu)$. In this simulation, we considered four models, all with $\beta_0 = -1.5$, $\beta_1 = -0.25$, $\beta_2 = 0.23$, $\gamma_0 = -6.0$, $\gamma_1 = 0.35$ and $\gamma_2 = -0.25$, but with different values for σ^2_ν: 0.2, 0.4, 0.6. In order to apply the Bayesian methodology developed in this paper, we assigned normal prior distributions for the location parameters, given by $\beta_k \sim N(0, 10^5)$, $\gamma_k \sim N(0, 10^2)$, $k = 0, 1, 2$, and a gamma prior distribution, for the variance of the random effect, $\sigma^2_\nu \sim Gamma(0.01, 0.01)$, as in the last section.

The posterior parameter estimates obtained applying the proposed MCMC algorithm and

WinBugs software are given in Table 4. The first row shows the parameter estimates when the true value of σ_ν^2 is $\sigma_\nu^2 = 0.2$, obtained by applying the proposed MCMC algorithm, developed using MathLab software (PA). The second row shows the parameter estimates obtained using WinBugs, and so on. The results presented in Table 4 are based on one long chain of size 10000 and burn-in period of size 3000. For all parameters, in all simulations, the parameter estimates are very close to the respective true values, all of them with small standard deviations.

Parameters		β_0	β_1	β_2	γ_0	γ_1	γ_2	σ_ν^2
$\sigma_\nu^2 = 0.2$	Mean	-1.503	-0.250	0.230	-6.234	0.398	-0.257	0.226
PA	s.d.	0.006	0.001	0.000	0.429	0.063	0.027	0.211
$\sigma_\nu^2 = 0.2$	Mean	-1.503	-0.250	0.230	-6.284	0.401	-0.254	0.231
WB	s.d.	0.006	0.000	0.000	0.431	0.059	0.028	0.338
$\sigma_\nu^2 = 0.4$	Mean	-1.506	-0.250	0.230	-6.251	0.364	-0.242	0.375
PA	s.d.	0.008	0.000	0.000	0.417	0.064	0.028	0.297
$\sigma_\nu^2 = 0.4$	Mean	-1.505	-0.2503	0.2304	-6.377	0.374	-0.2379	0.5013
WB	s.d.	0.008	0.000	0.001	0.416	0.067	0.0295	0.368
$\sigma_\nu^2 = 0.6$	Mean	-1.505	-0.250	0.230	-6.315	0.468	-0.269	0.817
PA	s.d.	0.004	0.001	0.000	0.531	0.062	0.036	0.523
$\sigma_\nu^2 = 0.6$	Mean	-1.504	-0.250	0.230	-6.604	0.480	-0.255	1.023
WB	s.d.	0.004	0.001	0.000	0.562	0.072	0.037	0.599

Table 4: Posterior mean and standard deviation of the parameters of the gamma random effect model.

Another simulation study, with the same two random explanatory variables X_1 and X_2, also was considered, generating 40 observed values. The values y_i, $i = 1, 2, \ldots, n$, of the variable of interest Y were generated from a gamma distribution with mean $\mu_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$ and variance $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_{2i} + \gamma_2 x_{3i} + \nu_i)$, where $\nu_i \sim N(0, \sigma_\nu^2)$. In this simulation, we considered four models, all with $\beta_0 = 3$, $\beta_1 = 1.5$, $\beta_2 = 3.5$, $\gamma_0 = -5$, $\gamma_1 = 0.35$ and $\gamma_2 = 0.15$, but with different values of σ_ν^2: 0.0, 0.2, 0.5, 0.8. To apply the Bayesian methodology, we assigned the same prior distributions given in section 4.1. As a result, the estimates of the mean parameters of the model with $\sigma_\nu^2 = 0.2$, given by the posterior means, and the respective standard deviations, are

$\hat{\beta}_0 = 2.939(0.07946)$, $\hat{\beta}_1 = 1.508(0.02159)$ and $\hat{\beta}_2 = (3.496)0.00847$. For the other models, with other values of the σ^2_ν, the estimates of the mean parameters are also very close to the true values and with very small standard deviations. The estimates of the variance parameters was similar to the last simulation for $\sigma^2_\nu = 0.0$, 0.2 and 0.4.

For the model with $\sigma^2_\nu = 0.8$, the random effect entails that a reduced model to be considered as the best. The model with all the parameters in the variance model has a DIC value of 37.685, and the posterior estimates of the variance parameters are given by $\hat{\gamma}_0 = -4.156(1.05)$, $\hat{\gamma}_1 = 0.2193(0.1375)$, $\hat{\gamma}_2 = 0.09309(0.06409)$ and $\hat{\sigma}^2_\nu = 1.417(1.353)$. For the model without X_1 in the variance model, the value of the DIC criterion is 35.207. The model without X_2 in the variance model has a DIC value of 37.723 and the model with $\sigma^2_i = \exp(\gamma_0 + \nu_i)$ a DIC value of 36.987. Finally, the DIC value for the model with constant variance is 53.173. Thus, the best model is the one that has $\sigma^2_i = \exp(\gamma_0 + \gamma_2 x_{2i} + \nu_i)$ as variance model.

4.3. Fit of small correlated normal data

In a final simulated study values of the interest variables were generated assigning small correlation between observations of the interest variable $\boldsymbol{Y}$. In the following two cases considered in this simulation, 10 values of the explanatory variable were generated from a uniform distribution defined in the interval $(0, 10)$ and values of $\boldsymbol{Y}$ were generated from a multivariate normal distribution with mean model given by $\mu_i = 2.5 + 3.5x_i$, $i = 1, 2, \ldots, 10$, and with variance-covariance matrix with diagonal elements given by $\sigma^2_{ii} = \exp(-2.5 + 0.45x)$ and small non diagonal elements: 0.03 in the first simulation and 0.1 in the second simulation.

Heteroscedastic linear normal regression models were fitted to this systematic data sets, with the aim of comparison with the result obtained applying models with random effects. In both cases, the normal prior distribution was given by $\beta_k \sim N(0, 10^5)$, $\gamma_k \sim N(0, 10^2)$, $k = 0, 1$, and the gamma prior distribution for the variance of the random effect that is $\sigma^2_\nu \sim Gam(0.01, 0.01)$, as in the last section.

The posterior parameter estimates for each of the proposed models are showed in Table 5. For each of the simulation, we can see that the normal RE model fit the proposed data set better than the heteroscedastic regression models. Therefore, we can conclude that the RE normal models are

First simulation. $Cov(Y_i, Y_j) = 0.03$, $i \neq j$						
	DIC	β_0	β_1	γ_0	γ_1	σ_e
Model with R.E.	25.072	2.242 (0.2856)	3.533 (0.1199)	-2.714 (1.607)	0.3183 (0.3536)	5.546 (15.05)
Model without R.E.	29.500	2.318 0.3369	3.538 0.1561	-2.329 1.173	0.403 0.1982	- -
Second simulation, $Cov(Y_i, Y_j) = 0.1$, $i \neq j$						
Model with R.E.	21.433	2.85 0.829	3.639 0.1375	-6.498 4.82	0.9932 0.6549	6.061 15.1
Model without R.E.	34.893	2.839 0.9416	3.588 0.1574	-3.801 2.903	0.6852 0.4175	- -

Table 5: Parameter estimates (small correlated normal data)

appropriate to analyze slightly correlated data set.

5. A real data set application

In this section, we deal with three different real data sets. The first one is a growth model to describe normal data; the second one is a model for non-negative data, described by a gamma distribution and the third one is a model for a rate obtained from educational data, modeled by a beta regression model.

5.1. Growth and development of babies: an application of normal random effect models.

In this section, we analyze the growth and development of some groups of babies from zero to thirty months of age. The data set was collected by Pediatricians in hospitals located in Bogotá, Colombia and by students of the Andes University from the beginning of 2000 to the end of 2002. The data set shows some interesting characteristics: (a) The height increases with time but it does not have a linear behavior, as shown in Figure 1. (b) Sample variance is not homogeneous, it seems to increase at an initially small time interval and then it decreases. As seen, the performance of the proposed algorithm is similar to that obtained using WinBugs. The data analysis discussed below is based on this software.

1. **Heteroscedastic linear regression models**

   From the behavior of the data (black dots in Figure 1), we can see that it follows the behavior of the square root function, translated over Y by the quantity $(0, \beta_0)$, where β_0 is a real number close to 50. Thus, taking into account the behavior of the variance, we propose to fit the data to the model

$$\boldsymbol{Y}_i = \beta_0 + \beta_1\sqrt{x_i} + \boldsymbol{\epsilon}_i,\ \ \boldsymbol{\epsilon}_i \sim N(0, \sigma_i^2), \tag{14}$$

   where the variance model is $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_i)$. In this model, $\boldsymbol{Y}_i$ denotes height of the i-th baby and x_i age. Although we consider other linear models to fit this data set, this model is the most appropriate to describe the behavior of this data set.

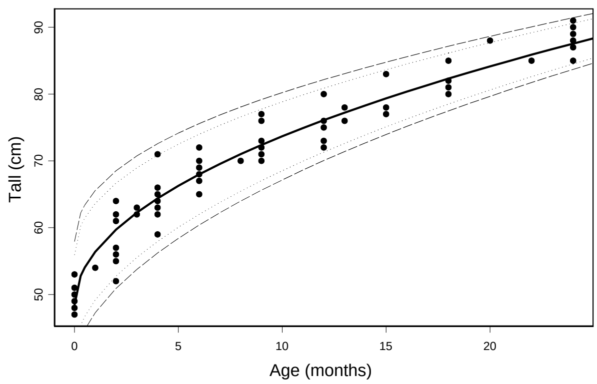

Figure 1: Prediction: The darker line represents the fitted posterior mean; the dotted lines represent the 90 % and 95 % prediction interval; the points correspond to the observations.

The posterior means and posterior standard deviations for the parameters in this model are given by: $\hat{\beta}_0 = 48.408(0.728)$, $\hat{\beta}_1 = 7.990(0.221)$, $\hat{\gamma}_0 = 2.303(0.2529)$, $\hat{\gamma}_1 = -0.038(0.022)$.

For this model, the DIC value is 356.088. When we consider model (14), but with constant variance, the DIC value is given by 356.798.

For the same data set, we propose model (14) for the mean and model $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 t_i + \nu_i)$ for the variance, where $\nu_i \sim N(0, \sigma_\nu^2)$. The posterior means and standard deviations for the model parameters are: $\hat{\beta}_0 = 48.460(0.759)$, $\hat{\beta}_1 = 7.966(0.233)$, $\hat{\gamma}_0 = 2.286(0.291)$, $\hat{\gamma}_1 = -0.03537(0.0248)$ and $\hat{\sigma}_\nu^2 = 0,07363(0,1476)$. The DIC value for this model is 356.221.

Finally, we consider the same model for the mean and $\sigma_i^2 = \exp(\nu_i)$ for the variance, where $\nu_i \sim N(0, \sigma_\nu^2)$. The posterior means and standard deviations for the model parameters are: $\hat{\beta}_0 = 48.95(0.376)$, $\hat{\beta}_1 = 7.516(0.1049)$ and $\hat{\sigma}_\nu^2 = 6.397(2.067)$. The DIC value for this model is 331.520.

For all models, the mean parameter estimates are very similar. However, based on the DIC values, the model with variance structure given by $\sigma_i^2 = \exp(\nu_i)$ gives better fit for the tallness data, since it has smaller DIC values.

Figure 1 shows the posterior mean height curve of babies and the posterior 90 % credibility interval. From Figure 1, we can see that the babies, all of whom were receiving nutritional supplement wealth, have heights within international parameters. Thus, there is no evidence of problems in the growth process. It is important to see that tallness is the most important parameter to assess the nutritional state and growth conditions of babies. For all models, the behavior of the chains corresponding to each parameter has a small transient stage, indicating the convergence speed of of the chain. Figure 2 shows the chain samples for the first 4500 iterations, for the first model. The other results reported in this section are based on a sample of 4000 draws after a burn-in of 1000 draws to eliminate the effect of initial values. The histogram of these samples shows that the posterior marginal distribution for all the parameters is approximately normal.

2. Heteroscedastic nonlinear regression models

Nonlinear regression is usually applied in many areas of science, such as biology, where the variable of interest is for example tallness, weight, enzyme activation, blood pressure or

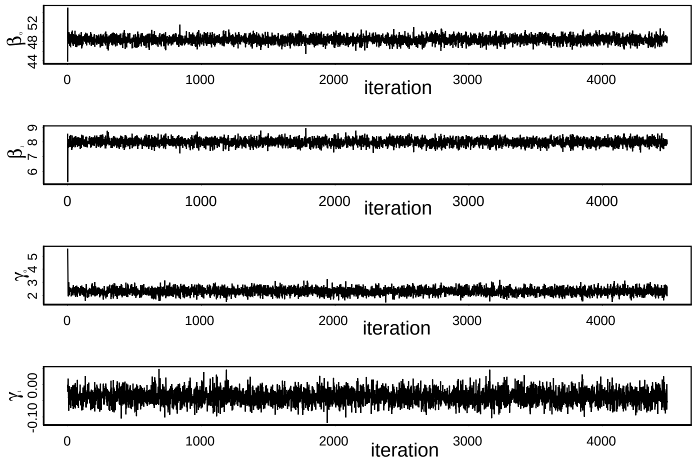

Figure 2: Behavior of the chain sample for parameters of the mean model β_i, and parameters of the variance model γ_i, i=0,1.

temperature. As examples, nonlinear models are used to fit radioligand and enzyme kinetics data or to fit dose response curves (Motulsky and Christopoulos, 2004). As a way to introduce a double generalized nonlinear model with random effects, we analyzed the growth and development data set considered in section 5.1, through the usual nonlinear model

$$\mu_i = \beta_0/(1 + \beta_1 \exp(\beta_2 x_i)) \tag{15}$$

for the mean and $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_i)$ for the variance. This model is appropriate to analyze this type of data, especially when it has asymptotical behavior (Cepeda-Cuervo and Núñez-Anton, 2009).

To get samples of the posterior distribution, the proposed distribution function is obtained with normal prior distribution and with the working observational model obtained by assigning a normal distribution to the first degree Taylor approximation of the nonlinear function (15).

The posterior means and standard deviations for the model parameters are: $\hat{\beta}_0 = 90.41(2.624)$, $\hat{\beta}_1 = 0.7213(0.0371)$, $\hat{\beta}_2 = -0.1197(0.016439)$, $\hat{\gamma}_0 = 2.46(0.3222)$, $\hat{\gamma}_1 = -0.02835(0.03059)$. For this model, the DIC value is 372.508

Next, we consider the same model for the mean, but with variance equal to $\sigma_i^2 = \exp(\gamma_0)$.

For this model, the posterior mean and standard deviation are $\beta_0 = 88.77(1.825)$, $\hat{\beta}_1 = 0.7071(0.03270)$, $\hat{\beta}_2 = -0.1303(0.01293)$, $\hat{\gamma}_0 = 2.2080(0.1713)$. For this model, the DIC value is 371.997.

Finally, we consider de same mean model but with the variance model given by $\sigma_i^2 = \exp(\nu_i)$, where $\nu_i \sim N(0, \sigma_\nu^2)$. In this case, the posterior mean and standard deviation are $\hat{\beta}_0 = 88.22(2.626)$, $\hat{\beta}_1 = 0.6862(0.04124)$, $\hat{\beta}_2 = -0.1168(0.01598)$, $\hat{\sigma}_\nu^2 = 7.165(2.562)$. The DIC value is 357.808. Thus, this model has the smallest DIC value among all the models of this type, and is the best.

5.2. Property tax data analysis: joint modeling of the mean and dispersion parameters in the gamma distribution.

A previously unpublished analysis of property tax (IPTU) data from Recife, Brazil, by Gauss Cordeiro and Enivaldo Rocha, professors at the Universidade Federal de Pernambuco, demonstrated that the IPTU tax has a gamma distribution and depends basically on the following dichotomous variables: pavement (PAV), water service (WA), illumination ($ILLU$), occupation (ORD) and location (LOC). $PAV = 1$ if the property is on a paved street and $PAV = 0$ otherwise; $WA = 1$ if the property has water service and $WA = 0$ otherwise; $ILLU = 1$ if the property is on an illuminated street and $ILLU = 0$ otherwise; $ORD = 1$ if the property is in a district with an ordered occupation project and $ORD = 0$ in any other case; $LOC = 1$ if the property is on a local street, which does not connect directly to other places in the city, and $LOC = 0$ in any other case.

To apply the Bayesian methodology, we considered a normal prior distribution, with large variance to express non prior information, as in the last example. In the analyzes of the property tax (IPTU) data, considering DGLM without random effects, we found the following model as the best:

$$\begin{aligned} \mu &= \beta_0 + \beta_1 PAV + \beta_2 WA + \beta_3 ILLU + \beta_4 ORD + \beta_5 LOC \\ \sigma^2 &= \exp(\gamma_0 + \gamma_1 PAV + \gamma_2 WA + \gamma_3 ORD). \end{aligned}$$

The posterior parameter estimates (and posterior standard deviations) are given by: $\hat{\beta}_0 = 699.100(30.4)$, $\hat{\beta}_1 = 393.5(91.73)$, $\hat{\beta}_3 = 60.62(20.54)$, $\hat{\beta}_4 = 45.91(20.69)$, $\hat{\beta}_5 = -313.70(35.5)$, $\hat{\gamma}_0 = 8.175(0.2873)$, $\hat{\gamma}_1 = 1.89(0.3664)$, $\hat{\gamma}_3 = 1958(0.312)$. For this model, the DIC value is 2817.810.

A second model is given by:

$$\begin{aligned} \mu &= \beta_0 + \beta_1 PAV + \beta_2 WA + \beta_3 ILLU + \beta_4 ORD + \beta_5 LOC \\ \sigma^2 &= \exp(\gamma_0 + \gamma_1 PAV + \gamma_2 WA + \gamma_3 ORD + \nu). \end{aligned}$$

where $\nu \sim N(0, \sigma_\nu^2)$, σ_ν^2 unknown. In this case, we also consider normal prior distributions for β and γ and a gamma(0.001,0.001) prior distribution for σ_ν^2. The posterior parameter estimates (and posterior standard deviations) are: $\hat{\beta}_0 = 700.2(28.56)$, $\hat{\beta}_1 = 386(83.29)$, $\hat{\beta}_3 = 39.64(17.15)$, $\hat{\beta}_4 = 35.86(18.63)$, $\beta_5 = -316.2(31.97)$, $\hat{\gamma}_0 = 7.999(0.3863)$, $\hat{\gamma}_1 = 2.382(0.4702)$, $\hat{\gamma}_3 = 1.318(0.4213)$, $\sigma_\nu^2 = 0.9795(0.2707)$. The DIC value is 2734.205. Thus, it is the best model to fit the property tax (IPTU) data.

In this case we also can see that DGLM-RE has lower DIC values than the corresponding DGLM, so the former model better fits the property tax data.

5.3. Analysis of school enrollment rate data: an application of joint modeling of the mean and precision parameter in the beta distribution

This section includes an analysis of the schooling rate in Colombia, for the years 1991 to 2003. This rate is determined by dividing the number of enrolled students between 5 and 19 years of age by the number of potential students in this ages range. This is a measurement of the percentage of the population of school age with access to the educational system.

The information of the number of people registered in the school system was obtained from the DANE (Colombian sensus bureau) for the years ranging from 1991 to 1998, and from the Education Ministry for the years 1999 to 2003. The rate data (and the corresponding years) are: 51.25 % (1991), 56.79 % (1992), 59.32 % (1993), 63.09 % (1994), 68.82 % (1995), 67.84 % (1996), 69.14 % (1997), 74.18 % (1998), 74.76 % (1999), 76.71 % (2000), 75.18 % (2001), 76.23 % (2002), 79.15 % (2003). In Figure 3, the black dots represent the rate data, the darker line represents the fitted posterior mean, and the dotted lines represent the 90 % credibility interval of the mean.

In this case, we consider that the data come from a beta distribution, $\boldsymbol{Y}_i \sim B(\alpha_i, \lambda_i)$, i=1,2, ...,n. Thus we propose a joint modeling of the mean and precision parameters, μ and $\phi = \alpha + \lambda$. For the mean, we propose a logistic model, $logit(\mu_i) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}$, or a nonlinear model, $\mu_i = \beta_0/(1 + \beta_1 \exp(\beta_2 x_i))$. In the analysis of this data set, we consider a non-linear model for the mean and many possible models for ϕ without a random effect. In this process, we observed that the model with $\alpha_i + \lambda_i = cte.$ is the one that best fits the data, since its DIC value is the smallest when compared to the others. Its parameter estimates (and standard deviation) are given by: $\hat{\beta}_0 = 0.8023(0.0201)$, $\hat{\beta}_1 = 0.7172(0.0579)$, $\hat{\beta}_2 = -0.2521(0.0408)$, $\hat{\gamma}_0 = 6.8222(0.4682)$. This model has a DIC value equal to -68.304.

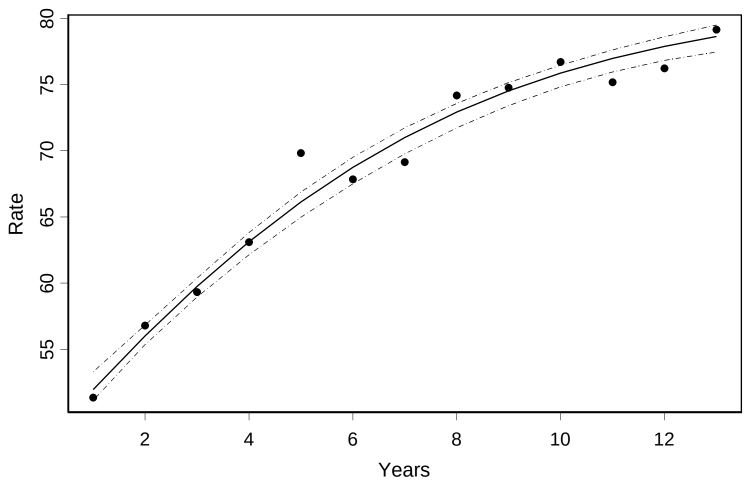

Figure 3: Mean rate of schooling data in Colombia.

Finally, we consider the nonlinear model for the mean and the model $a_i + b_i = \exp(\gamma_0 + \nu)$, where $\nu \sim N(0, \sigma_\nu^2)$, for τ. The posterior parameter estimations are $\hat{\beta}_0 = 0.8051(0.02407)$, $\hat{\beta}_1 = 0.7189(0.05826)$, $\hat{\beta}_2 = -0.2481(0.04459)$, $\hat{\gamma}_0 = 6.828(0.5222)$ and $\sigma_\nu^2 = 0.3684(1.175)$. For this model, the DIC value is -68.500, only a little smaller than the DIC value for the last model.

6. Conclusions

In this paper, the class of double stochastic generalized linear models is defined. This class of models includes double generalized models (Cepeda and Gamerman, 2005) and generalized linear models (McCullagh, 1989). The proposed models are fitted applying two Bayesian approaches. The first one is an extension of the general Bayesian methodology introduced by Cepeda and Gamerman (2005) . The second one is based on the use of the WinBugs software. The proposed models fit simulated and real data sets better than using double generalized linear models, especially when the observations of the variable of interest are weakly correlated. These models can be applied to data sets from many research areas, given that the Bayesian methodology proposed by Cepeda and Gamerman (2005) and extended in this paper can be easily implemented. The use of the WinBugs software also is a good alternative. The use of normal random effects in classes of models, such as EGARCH type models or Poisson process models, gives great flexibility of fit for the the data analysis, since one can find the extra-variability and the covariances between observations not depending on complex proposed parameterized models. Other distributions and different structures could be assumed for the random effects with the use of the proposed methodology.

7. Acknowledgments

Cepeda's work was supported by a grant from the Research Division of the National University of Colombia (Universidad Nacional de Colombia) and Migon's and Achcar's was partially supported by CNPq-Brazil and CAPES-Brazil grants. The authors are very grateful to Mariana Albi by a careful reading of this paper and her suggestions.

References

[1] Aitkin, M. (1987). Modeling Variance Heterogeneity in Normal Regression Using GLIM. *Applied Statistics*, **36**, 332-339.

[2] Hinde J. and Demetrio C.G.B.(1998). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151-170.

[3] Chip S. and Greenberg E. (1995). Understandingt he Metropolis-Hastingsa algorithm, . *The American Statistician*, **49**, 327-335.

[4] Cepeda, E.C. (2001). Variability Modeling in Generalized Linear Models, *Unpublished Ph.D. Thesis. Mathematics Institute, Universidade Federal do Rio de Janeiro.*

[5] Cepeda, C. E. and Gamerman, D. (2001). Bayesian Modeling of Variance Heterogeneity in Normal Regression Models. *Brazilian Journal of Probability and Statistics* , **14**, 207-221.

[6] Cepeda, E.C. and Gamerman, D. (2005). Bayesian Methodology for Modeling Parameters in the Two Parameter Exponential Family. *Revista Estadística*, **57(168-169)**, 93-105.

[7] Cepeda, C. E. and Achcar, J. (2010). Heteroscedastic Nonlinear Regression Models, *Communications in Statistics-Simulation and Computation,* **39**, 405-419.

[8] Cepeda, C. E. and Núnez-Anton, V. (2009). Bayesian Modelling of the Mean and Covariance matrix in Normal Nonlinear Models, *Journal of Statistical Computation and Simulation,* **79**, 837-853.

[9] Dey, D. K., Gelfand, A. E. and Peng, F. (1997). Overdispersed Generalized Linear Models, *Journal of Statistical Planning and Inference*, **64**, 93-107.

[10] Ferrari S, Cribari-Nieto F. (2004). Beta regression for modelling rates and proportions, *J Appl Stat.* **31**(7):799-815.

[11] Gamerman, D. (1997). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, *London: Chapman & Hall.*

[12] Gelfand, A.E. and Dalal, S.R. (1990). A Note on Overdispersed Exponential Families, *Biometrika*, **77**, 55-64.

[13] Hinde, J. and Demétrio, C. G. B. (1998). Overdisper-sion: Models and estimation. *Comput. Statist. Data Anal.*, **27**, 151-170.

[14] Jorgensen, B. (1997). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics*, **11**, 89-140.

[15] McCullagh, P. and J.A. Nelder. (1989). Generalized Linear Models. 2nd ed. Chapman & Hall/CRC, Boca Raton, Florida.

[16] Nelder, J. A. and Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments, *Appl.Stochast. Mod. Data Anal.*, **7**, 107-120.

[17] Quintero-Sarmiento A., Cepeda-Cuervo E. and Nunez-Anton V. (2011).Estimating infant mortality in Colombia: some overdispersion modelling approaches, *Journal of Applied Statistics*. To appear.

[18] Simas A. B., Barreto-Souza W., Rocha A. V. (2010). Improved Estimators for a General Class of Beta Regression Models, *Computational Statistics & Data Analysis*, **54**(2), 348-366.

[19] Smithson, M., and Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables, *Psychological Methods*, **11**(7), 54-71.

[20] Spiegelhalter. D.J., Best, N.G.: Carlin, B.P. and van der Linde. A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, (4),583-639.

[21] Williams D.A. (1982). Extra-binomial variation in logistic linear models, *Applied Statistics* **31**(2), 144-188.