Bayesian Generalized Linear Mixed Effects Models Using Normal-Independent Distributions: Formulation and Applications

Adeniyi, Isaac Adeola

Department of Mathematical Sciences, Federal University Lokoja, Nigeria, Department of Statistics, University of Ilorin, Ilorin, Nigeria

18 March 2020

Bayesian Generalized Linear Mixed Effects Models Using Normal-Independent Distributions: Formulation And Applications

Isaac Adeola Adeniyi[1]; **Waheed Babatunde Yahya**[2]

[1]Department of Mathematical Sciences, Federal University Lokoja, Nigeria.
Email: isaac.adeniyi@fulokoja.edu.ng

[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria.
Email: wbyahya@unilorin.edu.ng

ABSTRACT

A standard assumption is that the random effects of Generalized Linear Mixed Effects Models (GLMMs) follow the normal distribution. However, this assumption has been found to be quite unrealistic and sometimes too restrictive as revealed in many real-life situations. A common case of departures from normality includes the presence of outliers leading to heavy-tailed distributed random effects. This work, therefore, aims to develop a robust GLMM framework by replacing the normality assumption on the random effects by the distributions belonging to the Normal-Independent (NI) class. The resulting models are called the Normal-Independent GLMM (NI-GLMM). The four special cases of the NI class considered in these models' formulations include the normal, Student-*t*, Slash and contaminated normal distributions. A full Bayesian technique was adopted for estimation and inference. A real-life data set on cotton bolls was used to demonstrate the performance of the proposed NI-GLMM methodology.

Keywords: Generalized Linear Mixed Effects Models, Normal-Independent class, Normal density, Student-t, Slash density, Bayesian Method.

1.0 INTRODUCTION

A widely used technique for modelling clustered non-normally distributed data such as binary, count, skewed or other data is the generalized linear mixed-effects model (GLMM) framework (Schall, 1991; Zeger and Karim, 1991; Breslow and Clayton, 1993), which accommodates correlated observations through the incorporation of random effects. The Generalized linear mixed-effects models (GLMMs) generalize other models such as the

Linear Mixed-Effects Models (LMMs) and even the Linear Models (LMs) for modelling clustered (e.g. longitudinal) data which are common in clinical trials and epidemiological studies of cancer and in some other diseases.

Generally, the normality of the random effects is a common assumption in GLMMs but it may, sometimes, be unrealistic and too restrictive, obscuring important features of between-subjects variation. The presence of outliers also may cause the distribution of the random effects to be heavy-tailed and, thus, prevent the random effects to be adequately represented using the normality assumption. However, ignoring the departure from normality may cause biases or misleading results (Agresti *et al.*, 2004; Ghosh *et al.*, 2007; Verbeke and Lesaffre, 1996).

Neuhaus et al. (1992) through simulation showed that when there is misspecification of the distribution of random effects in a random-intercept logistic model, the Maximum Likihood Estimates (MLEs) of the model parameters for the fixed effects are inconsistent, but the magnitude of the bias is not large. However, estimates of the variance of the random effects exhibit large biases.

Heagerty and Kurland (2001) used the Kullback-Leible r Information Criterion to evaluate the consistency of MLEs of model parameters on conditional and marginal mean models. The authors showed that for conditionally specified models, misspecification of the random effects distribution may lead to seriously biased estimators for a cluster-level (between-subject) parameter and the intercept term when the variance of the random effects distribution is large.

Agresti *et al.* (2004) showed that the MLEs for fixed effect and variance component of the random effects distribution appear inconsistent when the true random effects distribution is a two-points mixture with a large variance in a simple one-way random-effects model.

Litiere et al. (2008), found that MLEs of between-subject parameters for the mean structure may be affected by misspecification of the random effects' distribution when the variance of the true random-effects distribution is large and estimates of the variance component are severely affected by misspecification in most situations.

Also, Litiere et al. (2007) studied the impact of the misspecification of the random effects distribution on the type I and type II error rates related to the Wald test for the mean structure parameters. They found that misspecification of the random effects distribution and the variance component of random effects can severely affect the power of the analysis and the type I error rate related to the tests for the intercept parameter.

To deal with the problem of wrongly specified random-effects distribution in LMM and Nonlinear Mixed Effects Models (NLMM), some proposals that have been provided involve replacing the assumption of normality by a class of elliptical distributions that cover both the light-and heavy-tailed distributions such as the Student- *t*, logistic and exponential power family or a class of Skew-elliptical distributions that include the multivariate skew-normal (SN) and skew-*t* (ST) distributions (Lin and Lee, 2007).

In Osiewalski and Steel (1993) and Osiewalski (1999), consideration was given to Bayesian approach to nonlinear models with elliptical distributions for the error term. Rosa *et al.* (2003) suggested the use of the normal-independent (NI) distributions (Liu, 1996) for LME models and adopted a Bayesian framework to obtain estimates. Savalli *et al.* (2006) and Osorio *et al.* (2007) studied LMMs using elliptical distributions while Lachos *et al.* (2011) studied LMMs using NI distributions.

Furthermore, robust modelling of Non-Linear Mixed-Effects Models (NLMEs) utilizing the normal-independent distributions can be found in Lachos *et al.* (2013) and Meza *et al.* (2012) while Chen and Luo (2016) proposed a Bayesian multilevel item response theory model using the normal-independent distributions.

In the case of GLMMs for clustered data, Chen *et al*. (2002) relaxed the normality assumption and required only that the distribution of random effects belong to a class of 'smooth' densities and approximate the density by the semi-nonparametric (SNP) approach of Gallant and Nychka (1987). In the study, a Monte Carlo EM algorithm using a rejection sampling scheme was used to estimate the fixed parameters of the linear predictor, variance components and the SNP density. However, many of the advantages of the parametric techniques do not easily carry over to the nonparametric setting (Samuels *et al*., 2012). Also, the method is only practically feasible for low dimensional random effects, and selecting the degree of the SNP polynomial is not a straight forward task.

Another issue with GLMMs and NLMMs is that the maximum likelihood estimates are obtained by integrating out the random effects which results to an integral without a closed-form and a non-linear maximization problem. If the random effects are low dimensional, then fitting via numerical integration using the Laplace approximations and adaptive quadrature can work well. However, the Laplace approximations can sometimes perform poorly such as in binary regression while adaptive quadrature may require a high computational burden to achieve high accuracy of the solutions and may not converge at all for complicated models (Zuur *et al.*, 2009). An alternative is to put priors on the parameters and use Markov Chain Monte Carlo (MCMC) sampling.

In this work, a class of generalized linear mixed-effects model where the assumption of normality is replaced by the class of NI distributions (Liu, 1996) that include the light- and heavy-tailed distributions is proposed. The NI distributions are an attractive class of symmetric heavy-tailed distributions that includes the normal distribution, the generalized Student-*t*, the Student-*t*, the Slash and the contaminated normal distributions as special cases. It is hoped that these distributions will provide an appealing robust alternative to the routine use of normal distributions in generalized mixed-effects models by allowing the random effects to have heavy tails and thus catering for random effects which are outliers.

Furthermore, we equally propose a full Bayesian estimation approach for estimating the parameters of the models. Although the NLMMs and LMMs with the NI distributions have appeared in the literature, to the best of our knowledge till now, there are no studies on Bayesian inference for GLMMs within the NI class. It is hoped that this approach eliminates the problems regarding departure from normality since the distributions in the NI class offer flexibility in shapes; easy implementations and applications under the Bayesian setting. The proposed NI-GLMMs modelling framework here is also tractable in the sense that they can preserve pleasant properties of other common distributions such that the parameters can be directly linked to some aspects of known probability density functions.

The rest of the paper is organized as follows; in section 2, a review of GLMMs, NI distributions and the Bayesian methodology are presented. The NI-GLMM framework, likelihood estimation are introduced in section 3. In section 4, we demonstrated the NI-GLMM methodology on a real-life dataset. Concluding remarks are given in Section 5.

2.0 REVIEW OF BACKGROUND TO METHODOLOGY

2.1 The Generalized Linear Mixed Effects Models

Clustered data are commonly encountered in practice. Examples of clustered data include split-plot designs in which the observations pertaining to the same block form a cluster and repeated measures data in which several observations are made sequentially on the same individual (cluster). Observations in the same cluster usually cannot be considered independent and mixed-effects models allow random effects that account for the cluster dependence and between-cluster variation (Pinheiro and Bates, 2000; Adeniyi *et al.*, 2018). In these models, the response is assumed to be a function of fixed (population) effects, non-observable cluster-specific random effects, and possibly an error term. Observations within the same cluster share common random effects and are therefore statistically dependent (Pinheiro and Bates, 2000).

In a mixed-effects model, the parameters can be divided into two classes: fixed effects which are associated with the average effect of predictors on the response, and variance-covariance components that are associated with the covariance structure of the random effects and of the error term (Pinheiro and Bates, 2000; Adeniyi *et al.*, 2018). The random effects are not considered as parameters but are commonly referred to as the Best Linear Unbiased Estimates of Predictors (BLUP) (Pinheiro and Bates, 2000).

A mixed-effects model in which both the fixed and the random effects contribute linearly to the response function is called the Linear Mixed-Effects Model (LMM) (Pinheiro, 1995). However, an LMM is not suitable for modelling a binary response, an ordinal response with few levels or a response that represents a count (Cameron and Trivedi, 1998). For these, we use the GLMMs.

In a generalized linear mixed model (GLMM) the conditional distribution of the response can be other distributions than the normal distribution. Common cases are the Bernoulli distribution for binary response data and the Poisson distribution for count response data. Because the expected value of each response may be restricted to an interval, (e.g. $(0, \infty)$ for the Poisson or $(0, 1)$ for the Bernoulli), the response is expressed usually as a non-linear function, $\boldsymbol{g}^{-1}$, called the inverse link function, of the linear predictor,

$$\eta = X^T\beta + Z^T b. \tag{1}$$

So,

$$g^{-1}(\eta) = g^{-1}(X^T\beta + Z^T b). \tag{2}$$

Let Y_i denote the vector of responses from subject (cluster) i, that is, $y_i = [y_{i1}, y_{i2}, \dots y_{in_i}]$. The marginal density of Y_i in the population is expressed as the following integral of the conditional likelihood $\ell(\cdot)$

$$h(y_i) = \int_{b_i} \ell\,(y_i|b_i) f(b_i) db_i \tag{3}$$

where

$$\ell\,(y_i|b_i) = \prod_{j=1}^{n_i} P\big(y_{ij}|b_i, X_{ij}, Z_{ij}\big)$$

and $f(b_i)$ represents the distribution of the random effects b_i, often assumed to be a multivariate normal density (Laird and Ware, 1982; Pinheiro and Bates, 2000). The marginal log-likelihood from the sample of m clusters is then obtained as

$$L(\boldsymbol{\beta}, \boldsymbol{D}|Y) = \sum_{i=1}^{m} log[h(y_i)]. \qquad (4)$$

Maximizing this log-likelihood yields maximum likelihood estimates for $\boldsymbol{\beta}$ and random effects covariance matrix $\boldsymbol{D}$.

Parameter estimation in GLMMs typically involves Maximum Likelihood (ML) or variants of ML techniques. Notwithstanding, the integrals in (4) above do not have closed-form expressions for non-normal GLMMs, hence, approximations to the integrals are used. Additionally, the solutions are usually iterative ones that can be relatively numerically intensive (McCulloch and Searle, 2001; Fahrmeir and Tutz, 2001).

2.2 The Normal-Independent (NI) Distributions

Following Lachos *et al*. (2013), a member of the NI family of distribution (Lange and Sinsheimer, 1993; Liu, 1996; Rosa *et al.*, 2003) is defined as the distribution of the k-variate random vector

$$Y = \mu + ZU^{-1/2}, \qquad (5)$$

where μ is a vector of location parameters, Z and U are independent such that Z is a normal random vector with mean vector 0, variance-covariance matrix Σ and U is a mixing nonnegative random variable with probability density function (pdf) $h(u|\nu)$ and cumulative distribution function (CDF) $H(u|\nu)$, where the scalar or parameter vector ν can be interpreted as a tail parameter which can be adjusted to absorb heavy tails.

It can be easily observed that given U, Y has a multivariate normal distribution with mean vector μ and variance-covariance matrix $u^{-1}\Sigma$. This implies that the NI distributions are scale mixtures of the normal distribution, where the distribution of U is the mixing distribution. Therefore, the pdf of Y is given by

$$NI(y|\mu, \Sigma, \boldsymbol{\nu}) = \int_0^{\infty} \phi_k(y; \mu, u^{-1}\Sigma) dH(u \mid \boldsymbol{\nu}),$$

where $\phi_k(.;\mu,\Sigma)$ stands for the pdf of the k–variate normal distribution with mean vector μ and variance-covariate matrix Σ. We use the notation $NI_k(\mu,\Sigma,H)$ when Y has distribution in the NI class. The three special cases we are considering within the NI class are the scale mixtures of multivariate normal distributions which include the multivariate Generalized Student-t, multivariate Slash, and multivariate contaminated normal distributions.

2.2.1 ***The multivariate generalized Student–t distribution***:

The multivariate generalized Student–t distribution denoted by $Gt_k(\mu,\Sigma,\nu_1,\nu_2)$, where ν is the degrees of freedom, is obtained from the mixture model (5) when U has the $Gamma(\nu_1/2,\nu_2/2)$ distribution, with $\nu_l > 0, l = 1,2$. The pdf of Y takes the following form:

$$GT(y|\mu,\Sigma,\nu) = \frac{\Gamma\left(\frac{p+\nu_1}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\pi^{\frac{k}{2}}} \nu_2^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \left(1+\frac{d}{\nu_2}\right)^{-\frac{k+\nu_1}{2}}, y \in \mathbb{R}_k, \quad (6)$$

where $\Gamma(\cdot)$ is the standard gamma function. Here, $\boldsymbol{\nu} = (\nu_1,\nu_2)$. Special cases of the generalized Student-t distribution are the Student-t distribution when $\nu_1 = \nu_2 = \nu$ and the Cauchy distribution, when $\nu_1 = \nu_2 = 1$. Also, when $\nu_1,\nu_2 \to \infty$, the normal distribution is obtained as the limiting case of (6).

2.2.2 ***The multivariate Slash distribution***:

The multivariate Slash distribution denoted by $SL_k(\mu,\Sigma,\nu)$ arises when the distribution of U in (5) is *Beta*(ν, 1), with $u \in (0, 1)$ and $\nu > 0$. Its pdf is given by

$$SL(y|\mu,\Sigma,\nu) = \nu \int_0^1 u^{\nu-1} \phi_k(y;\mu,u^{-1}\Sigma)du, \quad y \in R_k \quad (7)$$

The slash distribution reduces to the normal distribution when $\nu \to \infty$.

2.2.3 ***The multivariate contaminated normal distribution***:

The multivariate contaminated normal distribution is defined by $CN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\upsilon,\gamma)$ where $\upsilon,\gamma \in (0,1)$. In this case, U in (5) is a discrete random variable taking one of the two states in $\{\gamma,1\}$ with probability function given by

$$h(u|\boldsymbol{\nu}) = \upsilon\mathbb{I}_{\{\gamma\}}(u) + (1-\upsilon)\mathbb{I}_{\{1\}}(u),$$

where $\boldsymbol{\nu} = (\upsilon,\gamma)$ and $\mathbb{I}_{\{\tau\}}(.)$ is the indicator function of the set $\{\tau\}$. The associated density is given by

$$CN(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) = \upsilon\phi_k(\boldsymbol{y};\mu,\gamma^{-1}\boldsymbol{\Sigma}) + (1-\upsilon)\phi_k(\boldsymbol{y};\mu,\boldsymbol{\Sigma}). \quad (8)$$

Parameter ν can be viewed as the fraction of the data which are outliers while γ may be seen as a scale factor. As $\gamma \to 1$, the contaminated normal distribution reduces to the normal distribution.

2.3 Bayesian Approach to Estimation and Inference

Let θ be the parameter of interest with prior distribution $\pi(\theta)$ and let the realizations of the observed sample Y depend on θ, the probability distribution of Y given θ is $f(Y|\theta)$. By the Bayes theorem, posterior probability distribution of θ given the observed data Y is

$$P(\theta|Y) = \frac{f(Y|\theta) \times \pi(\theta)}{\int_\theta f(Y|\theta) \times \pi(\theta)\, d\theta} \tag{9}$$

The fundamental principle is that the posterior probability distribution of θ given observed data Y is only a function of the likelihood function and the prior distribution. However, in Bayesian estimation, closed-form or analytical expressions of (9) are often not available since the integrations involved are often of high dimensions and intractable except in few special cases.

Approximations such as the Laplace approximation are sometimes adopted to evaluate the integrals. Modern approaches include the Markov Chain Monte Carlo (MCMC) (Hastings, 1970; Geman and Geman, 1984; Gelman and Rubin 1992; Gelman *et al.* 2004; Brooks, 1998; Casella and George 1992; Gilks *et al.*, 1996) and the Integrated Nested Laplace Approximation (INLA; Li *et al.*, 2012; Rue *et al.*, 2009). The MCMC is now widely used because of its flexibility in the implementation and availability of high computing power. The Gibbs sampling (Geman and Geman, 1984), and the Metropolis-Hastings algorithm (Hastings, 1970) are common MCMC methods used to obtain Bayesian estimates as well as inference.

Inference and model selection under the Bayesian approach is usually done using credible intervals (Spiegelhalter *et al.*, 2004) and the Deviance Information Criterion (DIC; Spiegelhalter, 2002). The credible intervals using MCMC are obtained by taking the sample quantiles. For example, the upper and lower bound for a 95% credible interval is the 97.5th and 2.5th sample quantiles respectively.

3.0 THE NI-GLMM AND BAYESIAN ESTIMATION TECHNIQUE

3.1 Model Formulation

Let y_{ij} denote the j^{th} response for the i^{th} cluster, $i = 1,\ldots,m$ and $j = 1,\ldots,n_i$. For each i, conditional on random effects $b_i(q \times 1)$, the distribution of y_{ij}, $j = 1,\ldots,n_i$ is assumed to belong to the exponential family of distributions denoted by $Exp\left(F\left(\theta_{ij},\phi\right)\right)$ with density

$$f\left(y_{ij}|b_i;\, \beta,\phi\right) = exp\, \left\{\frac{y_{ij}\theta_{ij} - d\left(\theta_{ij}\right)}{a(\phi)} + c\left(y_{ij},\phi\right)\right\} \tag{10}$$

where $\mu_{ij} = E(y_{ij}|b_i) = d'(\theta_{ij})$ is the mean of Y_{ij}; ϕ is a dispersion parameter whose value may be known; $c(\cdot;\cdot)$ and $a(.)$ are arbitrary functions. The linear predictor (for GLMM) $\eta_{ij} = X_{ij}^T\beta + Z_{ij}^T b_i = g\left(\mu_{ij}^b\right)$ depends on fixed effects $\beta(p \times 1)$, the random effects b_i , and known vectors of covariates $X_{ij}(p \times 1)$ and Z_{ij} $(q \times 1)$ for the fixed and random effects, respectively. Also, b_i follows a continuous distribution with known density function. For example, in modelling clustered count data using the log link, the relationship between the mean of y_{ij}, $E(y_{ij}|b_i)$ and the set of covariates is given as $\log\left(\mu_{ij}\right) = X_{ij}^T\beta + Z_{ij}^T b_i$. Examples of distributions that can be expressed in terms of (10) include the binomial, Poisson, Normal, Gamma and the COM-Poisson (Adeniyi *et al.*, 2019; Conway & Maxwell, 1962; Shmueli et al., 2005) distributions.

Here, rather than make the usual assumption that b_i is standard multivariate normal, b_i is instead assumed to have a density in the NI class of densities described in Section 2.2 to account for possible departure from normality and allow the distribution of the random effects to be heavy-tailed. Therefore, b_i follows the $SL_p(\mu,\Sigma,\nu)$, $CN_p(\mu,\Sigma,\nu,\gamma)$ or the $Gt_p(\mu,\Sigma,\nu_1,\nu_2)$ in the proposed NI-GLMM.

Thus, a Normal-Independent generalized linear mixed-effects model (NI-GLMM) can be expressed as:

$$E\left(y_{ij}|\beta,b_i\right) = g^{-1}\left(X_{ij}^T\beta + Z_{ij}^T b_i\right)$$

$$b_i \overset{ind.}{\sim} NI_q(\mathbf{0},D,H),\ \ i = 1,\ldots,n.$$

The model is, therefore, written hierarchically as follows;

$$y_{ij}|b_i \sim Exp\left(F(\theta_{ij},\phi_{ij})\right) \tag{11}$$

$$b_i|U_i = u_i \sim N_q(\mathbf{0},u_i^{-1}\boldsymbol{D}), \tag{12}$$

$$U_i \sim H(u_i|\boldsymbol{\nu}), \qquad i = 1, \ldots, n. \tag{13}$$

where, $\mu_{ij} = E(y_{ij}|\beta, b_i) = d'(\theta_{ij}) = g^{-1}(X_{ij}^T\beta + Z_{ij}^T b_i)$, that is, $g(\mu_{ij}) = X_{ij}^T\beta + Z_{ij}^T b_i$; ϕ_{ij} is a dispersion parameter whose value may be known.

For example, if $y_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, is assumed to follow a Poisson distribution with the log link function and it is assumed that $b_i \sim SL_q(\mu, D, \nu)$, the hierarchical representation of the model is

$$y_{ij}|b_i \sim Pois\left(exp(X_{ij}^T\beta + Z_{ij}^T b_i)\right) \tag{14}$$

$$b_i|U_i = u_i \overset{ind.}{\sim} N_q(0, u_i^{-1}D) \tag{15}$$

$$U_i \sim Beta(\nu, 1), \qquad i = 1, \ldots, n. \tag{16}$$

3.2 The Likelihood and Estimation

3.2.1 *The Likelihood*

The assumption that a cluster's responses are independent given the random effects (and therefore can be multiplied to yield the conditional probability of the response vector) is known as the conditional independence assumption. Hence, a response variable y_i having cluster i with n_i units per cluster has the following conditional probability density function.

$$f\,(y_i|b_i) = \prod_{j=1}^{n_i} [f(y_{ij}|\, b_i)] \tag{17}$$

Let α be a vector of the distinct parameters of the random effects covariance matrix.

So,

$$f\,(y_i|b_i) = \prod_{j=1}^{n_i} \left[exp\ \left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\}\right]$$

The joint distribution of Y_i and b_i is

$$p\,(y_i, b_i) = p\,(y_i|b_i) \times f(b_i|\alpha)$$

The marginal density of y_i in the population is expressed as

$$p\,(y_i) = \int_{b_i} p\,(y_i, b_i) db_i = \int_{b_i} p\,(y_i|b_i) f(b_i|\alpha) db_i$$

So,

$$p\,(y_i) = \int_{b_i} \prod_{j=1}^{n_i} \left[exp\ \left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\}\right] f(b_i|\alpha) db_i \tag{18}$$

Let ϑ be the set of all parameters to be estimated. For a NI-GLMM $\vartheta = \{\beta, \phi, \alpha, \nu, \gamma\}$. The marginal likelihood from the sample of n clusters is

$$L(\vartheta; y) = \prod_{i=1}^{n} p(y_i)$$

Thus,

$$L(\vartheta; y) = \prod_{i=1}^{n} \left[\int_{b_i} \prod_{j=1}^{n_i} \left[exp \ \left\{ \frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\} \right] f(b_i|\alpha) db_i \right]$$

The marginal log-likelihood from the sample of *N* subjects is

$$\ell(\vartheta; y) = \log[L(\vartheta; y)] = \log \left[\prod_{i=1}^{n} p(Y_i) \right]$$

Hence,

$$\ell(\vartheta; y) = \log \left[\prod_{i=1}^{n} \left[\int_{b_i} \prod_{j=1}^{n_i} \left[exp \ \left\{ \frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\} \right] f(b_i|\alpha) db_i \right] \right]$$

$$\ell(\vartheta; y) = \sum_{i=1}^{n} \log \left[\int_{b_i} \prod_{j=1}^{n_i} \left[exp \ \left\{ \frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\} \right] f(b_i|\alpha) db_i \right] \quad (19)$$

Here, $f(b_i|\alpha)$ is a density of one of the $SL_p(\mu, \Sigma_\alpha, \nu)$, $CN_p(\boldsymbol{\mu}, D_\alpha, \nu, \gamma)$ or $Gt_p(\mu, D_\alpha, \nu_1, \nu_2)$ distributions. The notation D_α indicates that D is parameterized by α. Maximizing this log-likelihood yields ML estimates for $\boldsymbol{\beta}$, $\boldsymbol{\nu}$, γ, δ, and random effects covariance parameters vector $\boldsymbol{\alpha}$. However, the integrals are intractable, therefore, there is a need to approximate the integral. In this work, a full Bayesian approach is adopted.

3.2.2. *The Bayesian Framework*

Let $Y = \{y_i, i = 1,2,3, \dots, n\}$ be the observed data; ϑ be the set of all parameters to be estimated and $\pi(\vartheta)$ be the joint distribution of the parameters, the joint posterior distribution for the parameters is

$$\pi(\vartheta|Y) = \frac{\prod_{i=1}^{n} [f(y_i|\vartheta)] \, \pi(\vartheta)}{\int_\theta \prod_{i=1}^{n} [f(y_i|\vartheta)] \, \pi(\vartheta) d\vartheta}.$$

where $f(y_i|\vartheta)$ is the probability distribution of the data given ϑ. For a NI-GLMM, $\vartheta = \{\beta, \phi, \alpha, \boldsymbol{\nu}\}$ and $\pi(\vartheta) = \pi(\beta)\pi(\phi)\pi(\alpha)\pi(\boldsymbol{\nu})$. It should be noted that $L(\vartheta; y) = \prod_{i=1}^{n} [f(y_i|\vartheta)]$. Therefore,

$\pi(\vartheta|Y) \propto L(\vartheta; y)\pi(\vartheta)$

$$= \prod_{i=1}^{n}\left[\int_{b_i}\prod_{j=1}^{n_i}\left[exp\ \left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\}\right] f(b_i|\alpha)db_i\right]\pi(\vartheta). \quad (20)$$

The joint posterior distribution of the parameters for a NI-GMM is

$\pi(\vartheta|Y)$

$$= \frac{\pi(\vartheta)\prod_{i=1}^{n}\left[\int_{b_i}\prod_{j=1}^{n_i}\left[exp\ \left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c\left(y_{ij}, \phi\right)\right\}\right]\int_0^{\infty}\phi_q(b_i; 0, u_i^{-1}D_\alpha)h(u_i|\boldsymbol{\nu})du\, db_i\right]}{\int_\theta \pi(\vartheta)\prod_{i=1}^{n}\left[\int_{b_i}\prod_{j=1}^{n_i}\left[exp\ \left\{\frac{y_{ij}\theta_{ij} - d(\theta_{ij})}{a(\phi)} + c\left(y_{ij}, \phi\right)\right\}\right]\int_0^{\infty}\phi_q(b_i; 0, u_i^{-1}D_\alpha)h(u_i|\boldsymbol{\nu})du\, db_i\right]d\vartheta}$$

The full conditional posterior distributions for a NI-GMM are given as follows

$$\pi(\beta|D_\alpha, \boldsymbol{\nu}, \phi, b, Y) \propto \prod_{i=1}^{n}\left[\prod_{j=1}^{n_i}\left[f(y_{ij}|\beta, b_i)\right]\right]\pi(\beta), \quad (21)$$

$$\pi(D_\alpha|\beta, \boldsymbol{\nu}, \phi, b, Y) \propto \prod_{i=1}^{n}\left[f(b_i|D_\alpha, \boldsymbol{\nu})\right]\pi(D_\alpha), \quad (22)$$

$$\pi(b|\beta, D_\alpha, \boldsymbol{\nu}, \phi, Y) \propto \prod_{i=1}^{n}\left[\left(\prod_{j=1}^{n_i}\left[f(y_{ij}|\beta, b_i)\right]\right)f(b_i|D_\alpha, \boldsymbol{\nu})\right], \quad (23)$$

$$\pi(\boldsymbol{\nu}|\beta, D_\alpha, \phi, b, Y) \propto \prod_{i=1}^{n}\left[f(b_i|D_\alpha, \boldsymbol{\nu})\right]\pi(\boldsymbol{\nu}). \quad (24)$$

The expressions in (21 – 24) can further be written as

$$\pi(b|\beta, D_\alpha, \boldsymbol{\nu}, \phi, u, Y) \propto \prod_{i=1}^{n}\left[\left(\prod_{j=1}^{n_i}\left[f(y_{ij}|\beta, b_i)\right]\right)\phi_q(b_i|0, u_i^{-1}D_\alpha)\right], \quad (25)$$

$$\pi(u_i|\beta, D_\alpha, \boldsymbol{\nu}, \phi, b, Y) \propto \left(\prod_{j=1}^{n_i}\left[f(y_{ij}|\beta, b_i)\right]\right)\phi_k(b_i|0, u_i^{-1}D_\alpha)h(u_i|\boldsymbol{\nu}), \quad (26)$$

$$\pi(\boldsymbol{\nu}|\beta, D_\alpha, \phi, b, Y) \propto \left[\prod_{i=1}^{n}\left[h(u_i|\boldsymbol{\nu})\right]\right]\pi(\boldsymbol{\nu}) \quad (27)$$

$$\pi(\phi|\beta, D_\alpha, \boldsymbol{\nu}, b, Y) \propto \prod_{i=1}^{n}\left[\prod_{j=1}^{n_i}\left[f(y_{ij}|\beta, b_i, \phi)\right]\right]\pi(\phi), \quad (28)$$

The forms of $\pi(u_i|\cdot)$ and $\pi(\boldsymbol{\nu}|\cdot)$ depends on the specific NI distribution adopted and also on the prior for $\boldsymbol{\nu}$**.**

The prior distributions for the parameters are specified as follows.

i. $\beta \sim N(\beta_0, \Lambda)$,
ii. $D_\alpha \sim IW(\Omega, v)$,
iii. $\delta \sim N(\mathbf{0}, \tau)$,
iv. $\nu \sim Exp(\nu_0), I(\nu > 2)$ for the student-t distribution.
v. $\nu \sim Gamma(a, b)$, where a and b are small positive values such that $b \ll a$ for the slash distribution,
vi. $\nu \sim Beta(\nu_0, \nu_1)$ and $\gamma \sim Beta(\rho_0, \rho_1)$ for the contaminated normal distribution,

where the mutually independent Normal (N), Inverse Gamma (IG), Exponential (Exp) and Inverse Wishart (IW) prior distributions are chosen to facilitate computations (Davidian and Giltinan, 1995). Truncating the exponential distribution in the interval (2,∞) ensures finite variance. The super-parameter matrix Λ and Ω can be assumed to be diagonal for convenient implementation. The form of the prior for the dispersion parameter depends on the particular distribution assumed for the response. For example, $\phi = 1$ when the response follows the Poisson distribution and $\phi \sim IG(a_\phi, b_\phi)$ when the response follows the Normal distribution.

Lemma 1: Poisson GLMM with Slash distributed random effects

Suppose random variable y_{ij} indexes the count responses from an empirical real life scenario. Traditionally, y_{ij} is assumed to follow the Poisson distribution and its corresponding GLMM with Slash distributed random effects is given by the expressions (14-16) in section 3.1. The corresponding posterior distribution is given as

$$\pi(\vartheta|Y) \propto \prod_{i=1}^{n}\left[\int_{b_i}\prod_{j=1}^{n_i}\left[\frac{\theta_{ij}^{y_{ij}}e^{-\theta_{ij}}}{y_{ij}!}\right] f(b_i|u_i,\alpha)f(u_i)db_i\right]\pi(\vartheta), \tag{29}$$

where $\theta_{ij} = exp\left(X_{ij}^T\beta + Z_{ij}^T b_i\right)$, $f(b_i|u_i,\alpha)$ and $f(u_i)$ are the density functions of $N_q(0, u_i^{-1}D_\alpha)$ and $Beta(\nu, 1)$ distributions respectively.

3.2.3 *Estimation: Computing the Posterior*

In general, the integrals in (21-24) are usually of high dimension and do not have any closed form. Therefore, there are no analytical expressions for the posterior distributions. Analytic approximations to the integral may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of ϑ based on the observed data. As an alternative, MCMC procedures can be used to draw samples based on the posterior distributions given by (21-24) using the Gibbs sampling along with the Metropolis-Hastings (M-H) algorithm.

Now the Gibbs sampler method can be used to generate samples from the posterior distribution $\pi(\vartheta|Y)$ for a NI-GMM using the following algorithm.

Starting with initial values $\left(\beta^{(0)}, D_{\alpha}^{(0)}, \nu^{(0)}, \phi^{(0)}, b^{(0)}\right)$, at k-th iteration

i. Sample $\beta^{(k)}$ from $\pi\left(\beta|D_{\alpha}^{(k-1)}, \nu^{(k-1)}, \phi^{(k-1)}, \beta_{*}^{(k-1)}, b^{(k-1)}, Y\right)$;

ii. Sample $D_{\alpha}^{(k)}$ from $\pi\left(D_{\alpha}|\beta^{(k-1)}, \nu^{(k-1)}, \phi^{(k-1)}, D_{\alpha*}^{(k-1)}, b^{(k-1)}, Y\right)$;

iii. Sample $\nu^{(k)}$ from $\pi\left(\nu|\beta^{(k-1)}, D_{\alpha}^{(k-1)}, \phi^{(k-1)}, \nu_{*}^{(k-1)}, b^{(k-1)}, Y\right)$;

iv. Sample $\phi^{(k)}$ from $\pi\left(\phi|\beta^{(k-1)}, D_{\alpha}^{(k-1)}, \nu^{(k-1)}, \phi_{*}^{(k-1)}, b^{(k-1)}, Y\right)$;

v. Sample $b^{(k)}$ from $\pi\left(\phi|\beta^{(k-1)}, D_{\alpha}^{(k-1)}, \nu^{(k-1)}, \phi^{(k-1)}, Y\right)$,

for k = 1, 2, 3,…

After a suitable burn-in period (taken to be 40,000), a sample of $(\beta, \phi, \alpha, \nu, b)$ is obtained from the posterior distribution $\pi(\vartheta|Y)$. Repeating this process many times, we can obtain many independent samples from the target posterior distribution. Then, we approximate the posterior means and variances by their corresponding sample means and sample variances based on the simulated samples, which are the approximate Bayesian estimates of the means and variance-covariances. The credible intervals for inference purposes are obtained by taking the sample quantiles. For example, the upper and lower bound for a 95% credible interval is the 97.5$^{\text{th}}$ and 2.5$^{\text{th}}$ sample quantiles respectively.

4.0 APPLICATION TO COTTON DATA

The data used in this section come from a greenhouse experiment conducted by da Silva et al (2012) to determine the effect of artificial defoliation on cotton plants at different growth stages. In the experiment, five defoliation levels (0, 25, 50, 75 and 100%) were considered and the observed number of bolls produced by the plants at five growth stages (vegetative, flower-bud, blossom, boll and boll-open) was observed. The experimental unit was a pot with two plants. The number of cotton bolls was recorded at each culture cycle. Zeviani et al. (2014) used the gamma-count distribution to analyse the data, although the correlation among plants in the same pot was not accounted for.

In this study, we demonstrated the performances of the selected three members of the NI-GLMMs family (T-GLMM, SL-GLMM, and the CN-GLMM) formulated to model the number of bolls as a function of defoliation and growth stage while considering pot as a clustering variable. We compared the performances of these three models with the classical N-GLMM where the random effects are assumed to be normally distributed.

Here, our goal is to examine the effect of defoliation on the number of bolls produced as well as how the number of bolls produced varies across different growth stages. Hence, we propose to fit the following model:

$$\log(\mu_{ij}) = b_i + \beta_0 + \beta_1 def_{ij} + \beta_2 I(\text{boll}_{ij}) + \beta_3 I(\text{boll} - \text{open}_{ij}) + \beta_4 I(\text{flowerbud}_{ij}) + \beta_5 I(\text{vegetative}_{ij}), \quad (30)$$

where μ_{ij} is the expected number of cotton bolls produced by the j^{th} $(j = 1,2)$ plant in pot i; b_i, $i \in (1, \dots, 125)$ is the random effect associated with pot i; $I(\cdot)$ is the indicator function; β_0 is the fixed effect intercept and β_k's, $k \in \{1, \dots, 5\}$ are the fixed effects associated with the level of artificial defoliation ($def \in \{\text{x}: 0 \leq \text{x} \leq 100\}$) and growth stages. It should be noted that the blossom stage growth is the reference category and $n = 125$ while $n_i = 2$ for $i \in (1, \dots, 125)$. The following vague priors were set:

$\beta_k \sim N(0{,}1000)$, $k = 0, \dots, 5$; $\sigma_b^{-2} \sim G(0.1{,}0.01)$, $\nu \sim Exp(3) I(\nu > 2)$ for the student-t distribution (T-GLMM), $\nu \sim Gamma(0.1{,}0.01)$, for the slash distribution (SL-GLMM), $\nu \sim Beta(1{,}1)$ and $\gamma \sim Beta(1{,}1)$ for the contaminated normal distribution (CN-GLMM).

Considering the above prior distributions, we set up two parallel independent runs of the Gibbs sampler chain with size 80,000 for each parameter, using the first 40,000 iterations as the burn-ins to eliminate the effect of the initial values and to avoid correlation problems, a thinning rate of 10 is considered, yielding a total sample of size 4,000.

The convergence of the MCMC chains was monitored using the trace plots and Gelman-Rubin $\hat{R}$ diagnostics (Gelman and Rubin, 1992). The $\hat{R}$ values as given in Table 1 and trace plots presented by Figures 2 – 5 indicate that the MCMC runs attained convergence. Also, we confirmed that the MCMC runs for the random effects b_i, $i = 1, \dots, 125$, also converged, though, for the sake of space, the trace plots for the random effects are not presented.

For the assessment of the fitted model, the values of the Expected Predictive Deviance (EPD; Chen and Huang, 2016; Huang et al., 2011) and the Residual Mean Squares (RMS; Chen and Huang, 2016; Huang et al., 2011) as obtained from each model are used to evaluate the predictive performance while the DIC (Spiegelhalter, 2002) is used to evaluate the goodness-of-fit. EPD is calculated by $EPD = E\left[\sum\left(y_{rep} - y_{obs}\right)^2\right]$, where the predictive value y_{rep} is a replicate of the observed y_{obs} and the expectation is taken over the posterior distribution of the model parameters ϑ (Gelman et al., 2003). The RMS for each MCMC chain is given by $\frac{\sum\left(y_{obs}-y_{fitted}\right)^2}{n_{obs}}$, where n_{obs} is the total number of observations which is $\sum n_i = 250$ in this case. Note that the reported results including RMS and EPD are based on the two MCMC chains.

Figure 1 represents plots of the number of cotton bolls recorded for each combination of the defoliation level and growth stage. The Bayes estimates of the fixed effects $\left(\hat{\beta}_k, k = 0, \ldots, 5\right)$, the random effects variance σ_b^2 with 95% Credible Interval (C.I.) given in brackets as well as model performance measures including the DIC, EPD and RMS values based on the four models under consideration are as given in Table 1. The estimates of the random effect parameter are given in the Appendix.

The Bayes estimates (posterior means), Standard Deviation (SD) and C.I. for the parameters are identical across the four models. The C.I. indicates that all the fixed effects parameters except β_2 are statistically significant at 5% level of significance since their corresponding Credible Intervals do not include the value zero. We also observe that defoliation has a significant negative effect $\left(\hat{\beta}_1 = -0.004\right)$ on the number of bolls that agree with the plot in Figure 1.

The Bayes estimates of ν for T-GLMM, SL-GLMM and CN-GLMM are 7.532, 3.373 and 0.402 respectively while the estimate for γ is 0.674. These values indicate that the random effects may be slightly heavy-tailed. The results also indicate that CN-GLMM produced the least RMS.

Table 1: Posterior means, posterior standard deviations (SD), and 95% credible intervals for the fixed effects, random effects variance, shape parameters and goodness of fit measures from the application of four NI-GLMMs to the Cotton- bolls data

Criterion		***N-GLMM****	***T-GLMM***	***SL-GLMM***	***CN-GLMM***
EPD		5.200	5.211	5.201	5.206
RMS		1.300	1.293	1.292	1.288
DIC		898.10	901.1	900.1	899.1
Parameter					
β_0	$\hat{\beta}_0$	1.318	1.316	1.316	1.319
	SD	0.091	0.091	0.090	0.093
	C.I	(1.133, 1.490)	(1.130, 1.490)	(1.129, 1.485)	(1.135, 1.504)
	$\hat{R}$	1.001	1.008	1.002	1.002
β_1	$\hat{\beta}_1$	-0.004	-0.004	-0.004	-0.004
	SD	0.001	0.001	0.001	0.001
	C.I	(-0.006, -0.002)	(-0.005, -0.002)	(-0.005, -0.002)	(-0.005, -0.002)
	$\hat{R}$	1.002	1.004	1.001	1.002
β_2	$\hat{\beta}_2$	0.062	0.062	0.060	0.060
	SD	0.111	0.114	0.112	0.115
	C.I	(-0.154, 0.281)	(-0.161, 0.285)	(-0.164, 0.280)	(-0.163, 0.285)
		1.001	1.009	1.001	1.001
β_3	$\hat{\beta}_3$	0.380	0.379	0.377	0.376
	SD	0.104	0.106	0.105	0.109
	C.I	(0.180, 0.587)	(0.176, 0.588)	(0.170, 0.582)	(0.158, 0.588)
	$\hat{R}$	1.001	1.006	1.001	1.001
β_4	$\hat{\beta}_4$	0.303	0.303	0.301	0.299
	SD	0.106	0.109	0.105	0.109
	C.I	(0.098, 0.514)	(0.089, 0.518)	(0.097, 0.510)	(0.084, 0.511)
		1.001	1.008	1.001	1.001
β_5	$\hat{\beta}_5$	0.284	0.282	0.281	0.278
	SD	0.106	0.108	0.107	0.111
	C.I	(0.077, 0.494)	(0.077, 0.495)	(0.073, 0.495)	(0.065, 0.491)
	$\hat{R}$	1.001	1.005	1.001	1.001
σ_b^2	$\hat{\sigma}_b^2$	0.006	0.010	0.005	0.006
	SD	0.003	0.05	0.003	0.003
	C.I	(0.002, 0.015)	(0.003, 0.022)	(0.002, 0.013)	(0.002, 0.014)
	$\hat{R}$	1.001	1.004	1.001	1.001
ν	$\hat{\nu}$	-	7.532	3.373	0.402
	SD	-	3.950	0.984	0.285
	C.I	-	(2.656, 17.661)	(1.789, 5.609)	(0.012, 0.953)
	$\hat{R}$	-	1.001	1.001	1.002
γ	$\hat{\gamma}$	-	-	-	0.674
	SD	-	-	-	0.233
	C.I	-	-	-	(0.161, 0.988)
	$\hat{R}$	-	-	-	1.001

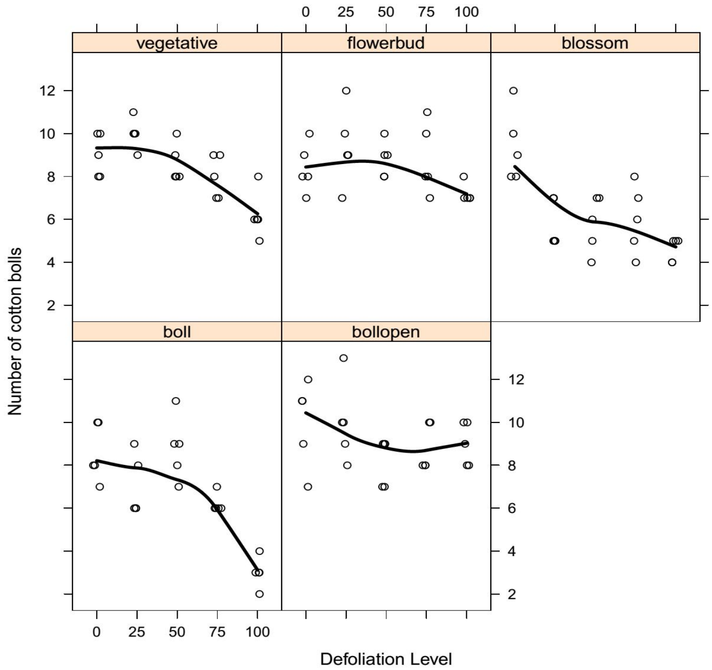

Figure 1: Number of cotton bolls against artificial defoliation level for each growth stage.

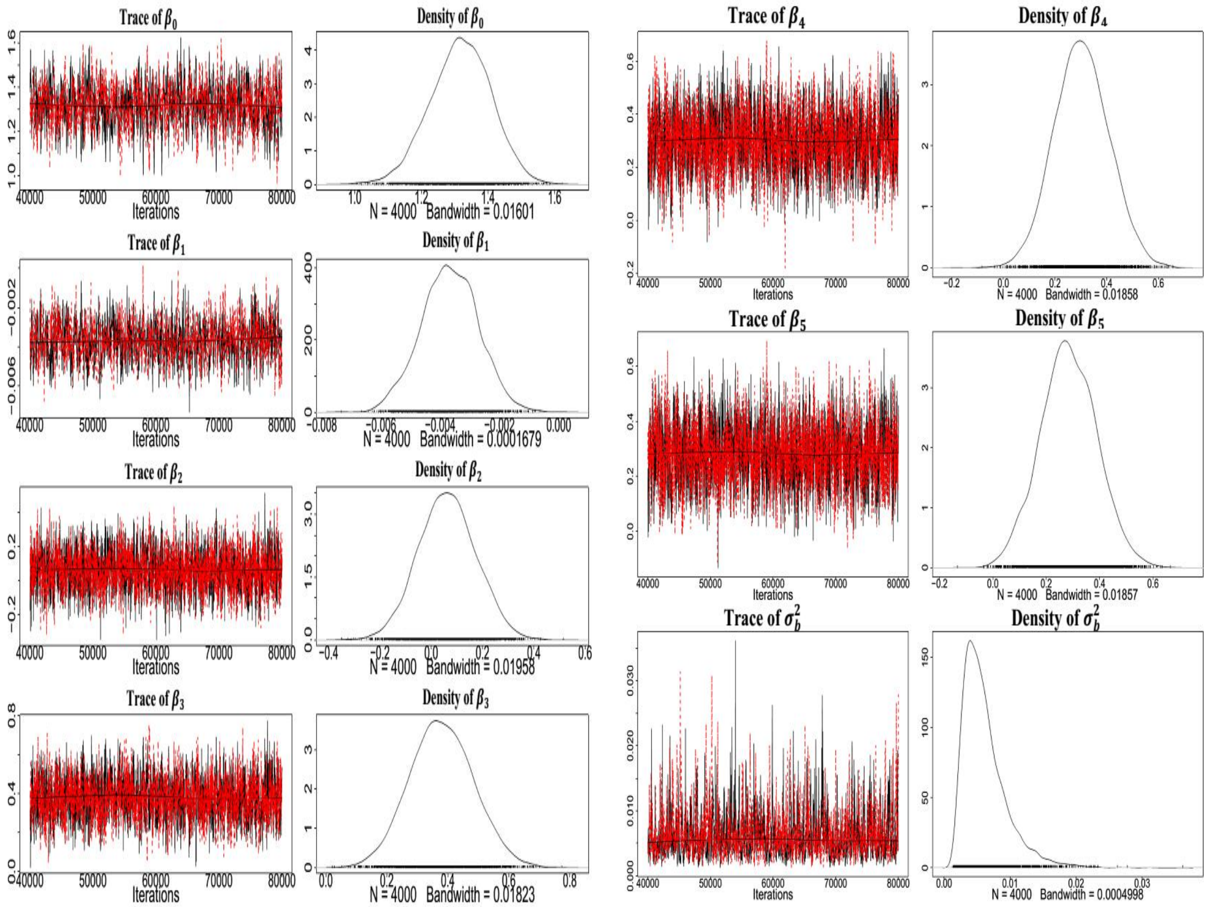

Figure 2: Trace and posterior density plots for MCMC samples of β_0, β_1 β_2, β_3, β_4, β_5 and σ_b^2 under the N-GLMM.

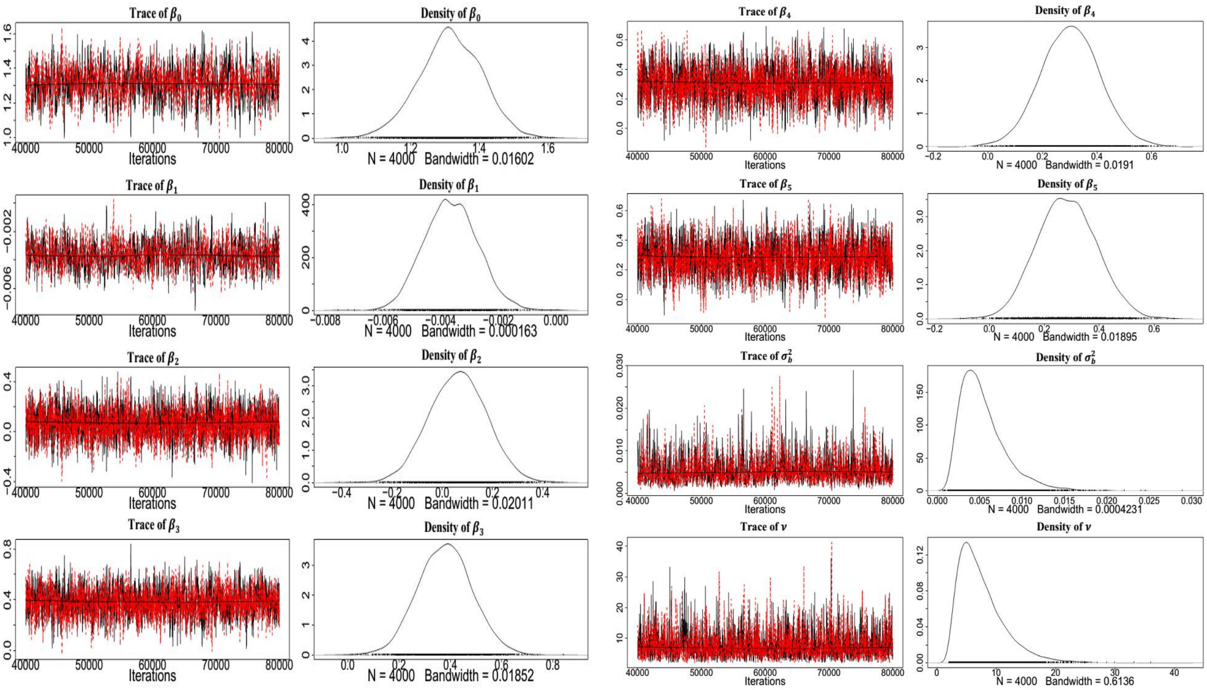

Figure 3: Trace and posterior density plots for MCMC samples of β_0, β_1, β_2, β_3, β_4, β_5, σ_b^2 and ν under the T-GLMM.

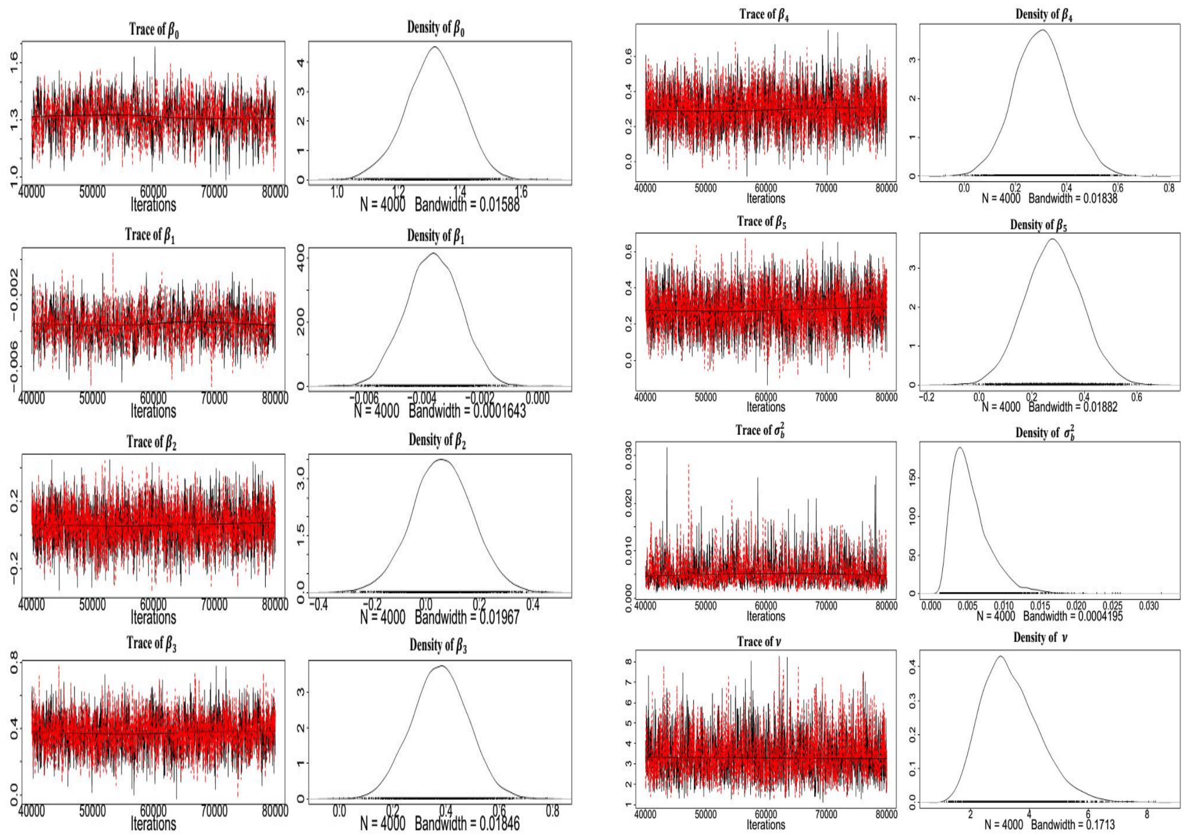

Figure 4: Trace and posterior density plots for MCMC samples of β_0, β_1, β_2, β_3 β_4, β_5, σ_b^2 and ν under the SL-GLMM.

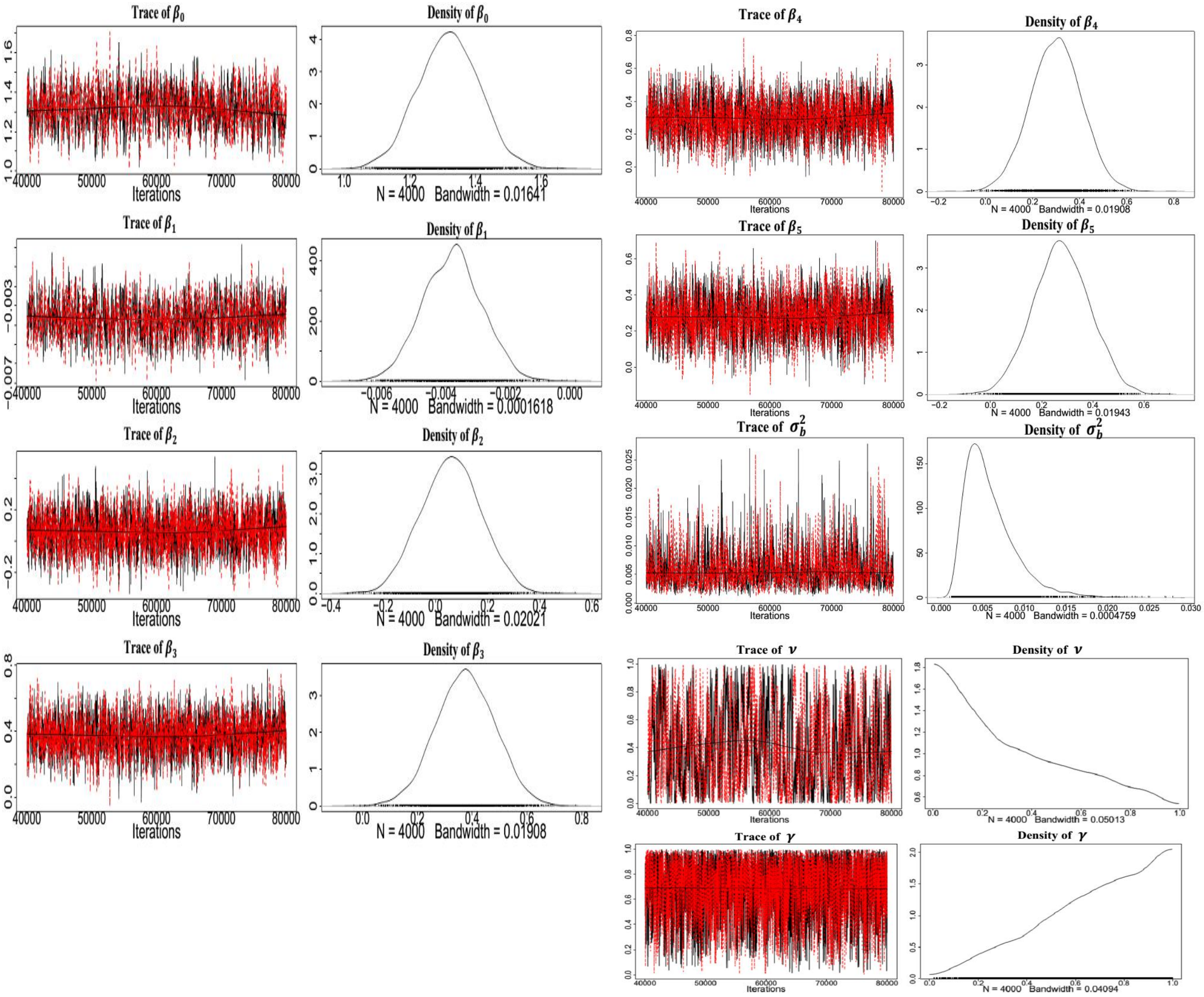

Figure 5: Trace and posterior density plots for MCMC samples of β_0, β_1, β_2, β_3 β_4, β_5, σ_b^2, γ and ν under the CN-GLMM.

5.0 CONCLUDING REMARKS

In this work, we discuss a Bayesian implementation of some robust generalized linear mixed-effects models using MCMC technique. The common assumption of normally distributed random effects terms is relaxed. Instead, the distribution of the random effects is allowed to belong to a class of flexible distributions known as the normal-independent (NI) distributions. The Student-*t*, the slash, and the contaminated normal distributions are the three particular cases considered in this study. These distributions which have thick tails are particularly robust to the presence of outliers. It is worthy of note that these three distributions generalizes the commonly used normal distribution.

A full Bayesian estimation technique using MCMC is adopted yielding a robust and flexible GLMM framework for modelling clustered data with non-normal responses. We call the proposed models and associated estimation procedure the NI-GLMM framework. Although the resulting posterior density functions are quite difficult to handle, we have shown that the estimation and inference can be carried out using MCMC methods. Also, we observed that the implementation of the technique is facilitated by the availability of stochastic representations of the distributions in the NI family.

Results obtained by application of the methodology to count data in an agricultural study on cotton plants provided a clear illustration of the implementation, flexibility and applicability of the proposed modelling framework. As can be observed from the results in Table 1 (including the one in the appendix) and Figures 1 to 5, the proposed NI-GLMM framework with better results is a useful alternative to the traditional GLMM where the normal distribution is usually assumed for the distribution of the random effects in the models.

It is quite instructive to remark that the appreciable performances of the proposed modelling framework in this study notwithstanding, this proposal is not the solution to all the modelling problems in the GLMMs. For instance, fitting he GLMMs with the use of the skew-elliptical distributions (Fernandez and Steel, 1998; Sahu *et al*., 2003; Azzalini *et al*., 1996; Azzalini *et al*., 1999) is already under consideration in our next study. We have equally conjectured the frequentist estimation and inference technique for the NI-GLMM presented here for a balanced comparison of the frequentist approach with the Bayesian method adopted in the current work.

Moreover, evaluation of the performance of the NI-GLMMs via extensive simulation studies has been carried out by the authors and will appear in the literature soon. Further applications of the NI-GLMM with binary response data or count data using flexible count distributions such as the COM-Poisson distribution (Adeniyi *et al.*, 2019; Conway & Maxwell, 1962; Shmueli et al., 2005) shall be presented in our subsequent works as well.

The codes for the implementation of the modelling procedure proposed in this work in R (R Core Team, 2019) and WinBUGS (Lunn *et al.*, 2000) are available from the authors upon request.

REFERENCES

Adeniyi, I.A., Shobanke, D. A. and Edogbanya H. O. (2019). Re-parameterization of the COM-Poisson Distribution Using Spectral Algorithms. *Pakistan Journal of Statistics and Operation Research*, 15(3): 701-712.

Adeniyi, I.A., Yahya, W. B. and Ezenweke C. P. (2018). A Note on Pharmacokinetics Modelling of Theophylline Concentration Data on Patients with Respiratory Diseases. *Turkiye Klinikleri Journal of Biostatistics*;10(1):27-45.

Agresti, A., Caffo, B. and Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *J. Comput. Graph. Statist.*, 47: 639–653.

Azzalini A. and Capitanio A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B*, 61: 579–602.

Azzalini A. and Dalla-Valle A. (1996). The multivariates skew-normal distribution. *Biometrika*, 8: 715–726.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association,* 88: 9–25.

Brooks, S. P. (1998) Markov chain Monte Carlo method and its application. *The Statistician*, 47: 69–100.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press. Cambridge, UK.

Casella, G. and George, E. (1992) Explaining the Gibbs sampler. *American Statistician*, 46: 167–74.

Chen, R. and Huang, Y. (2016). Mixed-Effects Models with Skewed Distributions for Time-Varying Decay Rate in HIV Dynamics. *Commun. Stat. Simul. Comput*, 45(2): 737–757.

Chen J., Zhang, D., and Davidian, M. (2002). A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics,* 3: 347–360.

Chen, G. and Luo, S. (2016). Robust Bayesian hierarchical model using normal/independent distributions. *Biometrical Journal,* 58(4): 831–85.

Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12, 132–136.

da Silva, A. M., Degrande, P. E., Fernandes, M. G., Suekane, R., and Zeviani, W. M. (2012). Impacto de diferentes níveis de desfolha artificial nos estádios fenológicos do algodoeiro. *Revista de Ciências Agrárias*, 35(1), 163-172.

Davidian, M. Gilitinan D.M. (1995) *Nonlinear models for repeated measurement data*. London: Chapman & Hall.

Fahrmeir, L. and Tutz, G.T. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, *2nd edition*. Springer-Verlag, New York.

Fernandez C. and Steel M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–371.

Gallant, A. R. and Nychka, D. W. (1987). Seminonparametric maximum likelihood estimation. *Econometrica,* 55: 363–390.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6: 721–741.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis,* Second Edition. New York: Chapman and Hall.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.

Ghosh P., Branco M., and Chakraborty H. (2007). Bivariate random effect model using skew normal distribution with application to HIV-RNA. *Statistics in Medicine*, 26: 1255–1267.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57: 97–109.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88: 973–985.

Huang, Y. Chen, R. and Dagne, G. (2011). Simultaneous Bayesian Inference for Linear, Nonlinear and Semiparametric Mixed-Effects Models with Skew-Normality and Measurement Errors in Covariates. *The International Journal of Biostatistics*, 7(1).

Lachos V. H., Bandyopadhyay D. and Dey D.K. (2011). Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. Biometrics. 2011; 67(4):1594–1604.

Lachos, V. H., Castro, L. M., and Dey, D. K. (2013). Bayesian inference in nonlinear mixed–effects models using normal independent distributions. *Computational Statistics & Data Analysis*, 64: 237–252.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38: 963-974.

Lange, K. L., Little, R., and Taylor, J. (1989). Robust statistical modeling using t distribution. *Journal of the American Statistical Association*, 84: 881–896.

Lange, K. L. and Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2: 175–198.

Li, Y., Brown, P., Rue, H., al Maini, M., and Fortin, P. (2012) Spatial modelling of lupus incidence over 40 years with changes in census areas. *Journal of the Royal Statistical Society: Series C*, 61 (1), 99-115.

Lin, T. I. and Lee, J. C. (2007). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**: 1490-1507.

Litiere, S., Alonso, A. and Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63: 1038–1044.

Litiere, S., Alonso, A. and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Stat. Med.*, 27: 3125–3144.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, 91.

McCullagh, P. and Nelder, J. A. (1997). *Generalized Linear Models, 2nd edition*. Chapman & Hall/CRC.

McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.

Meza, C., Osorio, F., and de la Cruz, R. (2012). Estimation in non-linear mixed effects models using heavy tailed distributions. *Statistics and Computing*, **22**: 121–139.

Neuhaus, J.M., Hauck, W.W. and Kalbfleisch, J.D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrik*a, 79: 755 – 762.

Osiewalski, J. (1999). Bayesian analysis of nonlinear regression with equicorrelated elliptical errors. *Test*, **8**: 339–344.

Osiewalski, J. and Steel, M. F. J. (1993). Robust Bayesian-inference in elliptic regression models. *Journal of Econometrics*, **57**, 345–363.

Osorio, F., Paula, G. A., and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**: 4354–4368.

Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*, 4: 12–35.

Pinheiro, J. and Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, New York.

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.Vienna, Austria. URL: http://www.R-project.org/.

Rodriguez, G. and Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 158: 73–89.

Rosa, G. J. M., Padovani, C. R., and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal*, 45: 573–590.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B*, 71 (2), 1-35.

Sahu S., Dey D. K., and Branco M. D. (2003). A new class of multivariate skew distribution with application to Bayesian regression models. *The Canadian Journal of Statistics*, 31(2): 129-150.

Samuels, M. L., Witmer, J. A., and Schaffner A. A. (2012). *Statistics for the Life Sciences*, 4th edition. Prentice Hall, Boston.

Savalli, C., Paula, G. A., and Cysneiros, F. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, **6**: 59–76.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika,* 78: 717-727.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S. and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, 54, 127–142.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approach to Clinical Trials and Health-care Evaluation*. John Wiley & Sons, Ltd, Chichester, UK.

Spiegelhalter, D. J. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 64(4): 583-639.

Verbeke, G. and Lesaffre, E. (1996). A liner mixed-effects models with heterogeneity in the random effects population. *Journal of the American Statistical Association*, 433: 217–221.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association,* 86: 79–86.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10, 325-337.

Zeviani, W. M., Riberio Jr, P. J., Bonat, W. H., Shimakura, S. E., and Muniz, J. A. (2014) The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*. 41, 2616-2626.

Appendix: The estimates of the Random Effect parameter (b) for the Cotton Bolls Data

Pot (i)	N-GLMM	T-GLMM	SL-GLMM	CN-GLMM
1	-0.0017	0.0010	0.0019	0.0005
2	-0.0058	-0.0053	-0.0063	-0.0060
3	-0.0123	-0.0148	-0.0159	-0.0138
4	-0.0109	-0.0125	-0.0143	-0.0134
5	0.0008	0.0012	0.0001	-0.0011
6	0.0093	0.0132	0.0131	0.0122
7	-0.0006	0.0013	0.0001	0.0008
8	0.0044	0.0070	0.0064	0.0063
9	0.0051	0.0072	0.0056	0.0066
10	0.0055	0.0072	0.0075	0.0071
11	-0.0017	-0.0023	-0.0023	-0.0026
12	-0.0009	-0.0014	-0.0024	-0.0008
13	0.0101	0.0122	0.0118	0.0114
14	-0.0023	-0.0035	-0.0015	-0.0021
15	0.0049	0.0042	0.0047	0.0048
16	0.0077	0.0091	0.0096	0.0101
17	-0.0034	-0.0035	-0.0061	-0.0041
18	-0.0029	-0.0034	-0.0043	-0.0046
19	0.0015	0.0019	0.0029	0.0037
20	0.0092	0.0107	0.0089	0.0093
21	0.0076	0.0079	0.0083	0.0068
22	-0.0039	-0.0085	-0.0077	-0.0052
23	-0.0070	-0.0065	-0.0064	-0.0057
24	-0.0128	-0.0133	-0.0118	-0.0130
25	-0.0045	-0.0061	-0.0060	-0.0069
26	-0.0183	-0.0204	-0.0227	-0.0212
27	-0.0120	-0.0147	-0.0149	-0.0158
28	-0.0127	-0.0157	-0.0156	-0.0143
29	-0.0072	-0.0093	-0.0102	-0.0063
30	-0.0006	0.0002	-0.0011	-0.0013
31	-0.0012	-0.0029	-0.0018	-0.0019
32	0.0156	0.0201	0.0192	0.0181
33	-0.0117	-0.0160	-0.0164	-0.0146
34	0.0039	0.0046	0.0050	0.0060
35	-0.0011	-0.0024	-0.0024	-0.0020
36	-0.0027	-0.0022	-0.0031	-0.0028
37	0.0019	0.0026	0.0042	0.0049
38	0.0026	0.0032	0.0034	0.0054
39	0.0086	0.0124	0.0113	0.0111
40	-0.0028	-0.0026	-0.0045	-0.0035
41	0.0183	0.0247	0.0248	0.0210

42	0.0134	0.0158	0.0172	0.0166
43	-0.0055	-0.0069	-0.0051	-0.0049
44	0.0009	0.0025	0.0019	-0.0001
45	0.0025	0.0007	0.0025	-0.0003
46	0.0009	-0.0003	0.0006	-0.0008
47	-0.0002	-0.0017	-0.0007	-0.0010
48	-0.0014	-0.0001	0.0015	0.0002
49	0.0000	-0.0009	0.0011	-0.0007
50	0.0059	0.0070	0.0064	0.0055
51	0.0135	0.0175	0.0199	0.0158
52	0.0090	0.0100	0.0108	0.0099
53	0.0037	0.0027	0.0027	0.0016
54	0.0249	0.0321	0.0352	0.0297
55	0.0020	0.0027	0.0026	0.0033
56	0.0007	0.0013	0.0024	0.0021
57	-0.0116	-0.0133	-0.0134	-0.0105
58	-0.0101	-0.0138	-0.0150	-0.0129
59	0.0027	0.0016	0.0000	0.0001
60	-0.0103	-0.0140	-0.0152	-0.0123
61	-0.0008	-0.0003	-0.0003	-0.0025
62	-0.0074	-0.0095	-0.0107	-0.0091
63	0.0072	0.0060	0.0036	0.0039
64	-0.0128	-0.0154	-0.0161	-0.0154
65	0.0056	0.0057	0.0052	0.0047
66	0.0130	0.0177	0.0167	0.0162
67	-0.0034	-0.0060	-0.0052	-0.0060
68	0.0075	0.0087	0.0103	0.0104
69	0.0010	0.0031	0.0032	0.0025
70	-0.0100	-0.0124	-0.0117	-0.0109
71	-0.0011	-0.0015	-0.0014	-0.0018
72	-0.0008	-0.0007	-0.0021	-0.0018
73	-0.0063	-0.0101	-0.0094	-0.0070
74	-0.0044	-0.0080	-0.0097	-0.0079
75	-0.0024	-0.0015	-0.0026	-0.0020
76	0.0010	-0.0006	-0.0005	0.0005
77	0.0124	0.0143	0.0147	0.0143
78	-0.0068	-0.0066	-0.0067	-0.0077
79	-0.0008	-0.0007	-0.0001	0.0000
80	0.0121	0.0137	0.0154	0.0137
81	0.0090	0.0127	0.0130	0.0111
82	-0.0060	-0.0101	-0.0090	-0.0088
83	-0.0079	-0.0109	-0.0106	-0.0093
84	0.0046	0.0050	0.0056	0.0047
85	-0.0088	-0.0093	-0.0094	-0.0085

86	0.0133	0.0179	0.0166	0.0149
87	0.0007	0.0026	0.0036	0.0020
88	0.0254	0.0335	0.0327	0.0289
89	0.0089	0.0102	0.0092	0.0084
90	0.0153	0.0184	0.0183	0.0152
91	-0.0008	-0.0010	-0.0008	0.0017
92	-0.0013	-0.0013	0.0003	0.0003
93	-0.0007	-0.0022	-0.0023	-0.0001
94	0.0000	-0.0021	0.0007	-0.0009
95	0.0040	0.0063	0.0082	0.0072
96	-0.0149	-0.0186	-0.0201	-0.0164
97	-0.0140	-0.0205	-0.0203	-0.0155
98	-0.0211	-0.0287	-0.0298	-0.0247
99	-0.0098	-0.0124	-0.0120	-0.0104
100	-0.0143	-0.0203	-0.0181	-0.0179
101	0.0001	-0.0018	-0.0011	-0.0006
102	-0.0234	-0.0279	-0.0286	-0.0230
103	-0.0098	-0.0135	-0.0148	-0.0124
104	0.0075	0.0078	0.0089	0.0070
105	-0.0003	0.0013	0.0000	-0.0004
106	-0.0052	-0.0081	-0.0063	-0.0086
107	0.0164	0.0187	0.0236	0.0200
108	-0.0121	-0.0147	-0.0138	-0.0135
109	-0.0009	0.0013	0.0002	-0.0007
110	-0.0011	-0.0021	-0.0012	0.0007
111	-0.0001	0.0013	-0.0007	-0.0017
112	-0.0130	-0.0148	-0.0154	-0.0145
113	-0.0123	-0.0147	-0.0154	-0.0137
114	0.0012	-0.0025	0.0003	0.0006
115	-0.0009	-0.0015	-0.0020	-0.0004
116	-0.0005	-0.0034	-0.0023	-0.0026
117	-0.0008	-0.0007	-0.0022	-0.0023
118	0.0097	0.0102	0.0119	0.0108
119	-0.0002	-0.0028	-0.0020	-0.0018
120	0.0110	0.0125	0.0125	0.0107
121	0.0072	0.0100	0.0108	0.0089
122	0.0018	0.0029	0.0028	0.0027
123	0.0147	0.0166	0.0180	0.0152
124	0.0011	0.0028	0.0038	0.0014
125	0.0151	0.0190	0.0168	0.0184