BAYESIAN ESTIMATION OF RANDOM PARAMETER MODELS OF RESPONSES WITH NORMAL AND SKEW-t DISTRIBUTIONS EVIDENCE FROM MONTE CARLO SIMULATION

Mohammad Masjkur[1] and Henk Folmer[2,3]

[1]**Department of Statistics**
Faculty of Mathematics and Natural Sciences
Bogor Agricultural University, Indonesia
[2]**Faculty of Spatial Sciences, University of Groningen, The Netherlands**
[3]**Department of Agricultural Economics**
College of Economics and Management
Northwest A&F University, Yangling, Shaanxi, China
masjkur@gmail.com, h.folmer@rug.nl

Abstract. Random parameter models have been found to outperform fixed parameter models to estimate dose-response relationships with independent errors. A major restriction, however, is that the responses are assumed to be normally and symmetrically distributed. The purpose of this paper is to analyze Bayesian inference of random parameter response models in the case of independent responses with normal and skewed, heavy-tailed distributions by way of Monte Carlo simulation. Three types of Bayesian estimators are considered: one applying a normal, symmetrical prior distribution, a second applying a Skew-normal prior and, a third applying a Skew-t-distribution. We use the relative bias *(RelBias)* and Root Mean Squared Error *(RMSE)* as valuation criteria. We consider the commonly applied linear Quadratic and the nonlinear Spillman-Mitscherlich dose-response models. One simulation examines the performance of the estimators in the case of independent, normally and symmetrically distributed responses; the other in the case of independent responses following a heavy-tailed, Skew-t-distribution. The main finding is that the estimator based on the Skew-t prior outperforms the alternative estimators applying the normal and Skew-normal prior for skewed, heavy-tailed data. For normal data, the Skew-t prior performs approximately equally well as the Skew-normal and the normal prior. Furthermore, it is more efficient than its alternatives. Overall, the Skew-t prior seems to be preferable to the normal and Skew-normal for dose-response modeling.

27

Abstrak. Model parameter acak diketahui lebih baik daripada model parameter tetap untuk menduga hubungan dosis-respons dengan galat acak bebas. Namun demikian, kendala utama adalah bahwa respons diasumsikan menyebar normal dan simetrik. Tujuan tulisan ini adalah menganalisa inferensia Bayesian dari model respons parameter acak pada kasus respons menyebar normal dan menjulur berekor panjang dengan metode simulasi Monte Carlo. Tiga tipe penduga Bayesian dipergunakan: pertama didasarkan pada sebaran prior normal, simetrik, kedua didasarkan pada sebaran-normal menjulur, dan ketiga didasarkan pada sebaran-t menjulur. Sebagai kriteria penilaian digunakan nilai bias relatif dan Akar Kuadrat Tengah Galat. Model yang digunakan adalah model linear Kuadratik dan model nonlinear Spillman-Mitscherlich. Simulasi pertama mengkaji unjuk kerja penduga pada kasus respons menyebar bebas, normal, dan simetrik; sedangkan yang lainnya pada kasus respons menyebar bebas, Skew-t berekor panjang. Temuan utama adalah bahwa penduga prior t-menjulur lebih baik daripada sebaran prior normal-menjulur dan normal simetrik pada data menjulur dan berekor panjang. Pada data normal, sebaran prior t-menjulur dapat disamakan dengan sebaran prior normal-menjulur dan prior normal. Sebaran prior t-menjulur lebih efisien dari keduanya. Secara umum, sebaran prior t-menjulur lebih baik daripada sebaran prior normal-menjulur dan normal bagi pemodelan dosis-respons.

Kata kunci: Model dosis-respons, penduga Bayesian, Gibbs sampler, model parameter acak, sebaran normal-menjulur, sebaran t-menjulur.

1. Introduction

The linear Quadratic and the nonlinear Spillman-Mitscherlich model are commonly applied to analyze dose-response relationship with independent errors in a large variety of fields including environmental sciences, biology, public health, and agricultural sciences (de Souza et al. [7]; Pinheiro et al. [23]; WHO [37]). The model parameters are usually estimated by means of least squares under the assumptions of fixed parameters and errors that are independently and normally distributed with constant variances (Lopez-Bellido et al. [17]; Sain and Jauregui [28]).

A limitation of the standard fixed parameter models is that they preclude the variability of the parameters that may exist among subjects. A model that does not have this limitation is the random parameter response model (Makowski and Wallach [19]; Makowski and Lavielle [20]; Plan et al. [24]; Tumusiime et al. [34]; Wallach [35]). This model type assumes that the response functions are common to all subjects, but that the parameters vary between subjects. For this purpose a random component is associated with the coefficients that represents inter-individual variability. The random parameter models have been found to

outperform the fixed parameter models (Boyer et al. [5]; Makowski et al. [18]; Makowski and Wallach [19]; Tumusiime et al. [34]).

The model parameters and the random errors are usually based on the assumption of independently, symmetrically, normally distributed response (Boyer et al. [5]; Makowski and Wallach [19]; Makowski and Lavielle [20]; Plan et al. [24]; Tumusiime et al. [34]). However, the assumption of normality may be too restrictive in many applications (Arellano-Valley et al. [1]-[2]; Jara et al. [11]; Ouedraogo and Brorsen [21]). Lachos et al. [14] proposed skewed linear mixed dose-response models when there is evidence of departure from symmetry or normality. The present paper deals with responses that follow the asymmetric heavy-tailed Skew-t distribution. The paper also considers responses that follow a normal distribution.

Random parameter dose-response models can be estimated by maximum likelihood (ML). However, for models that are nonlinear in the parameters, ML may lead to non-unique solutions (Brorsen [6]; Tembo et al. [33]). In addition, convergence may be problematic even with careful scaling and good starting values. Alternatively, Bayesian methods for which convergence of nonlinear estimation is not an issue, can be used (Brorsen [6]; Ouedraogo and Brorsen [21]). An additional advantage of the use of Bayesian methods is that the results are valid in small samples, which are quite common in dose-response modeling.

The objective of this paper is to investigate the performance of the Bayesian estimator with Skew-t prior of the random parameters linear Quadratic and the nonlinear Spillman-Mitscherlich model when the response follows (i) an asymmetric heavy-tailed Skew-t distribution, (ii) normal distribution. In addition to the Skew-t prior, we will also consider the commonly used normal, symmetrical prior, and the Skew normal prior.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce the Skew-Normal (SN) and Skew-t (S-t) distribution and specify the linear Quadratic and nonlinear Spillman-Mitscherlich model which in practice are most frequently used to model dose-response relationships. Section 3 presents the Bayesian inference approach as well as the model comparison criteria. Section 4 outlines the simulation framework and Section 5 the simulation results. Conclusions follow in Section 6.

2. The Skew SN And Skew S-t Distribution and the Linear Quadratic and Nonlinear Spillman-Mitscherlich Model

2.1. The Independent Skew-Normal (SN) and Skew-t (S-t) Distribution.
Lachos et al. [14] defined the family of Skew normal (SN) distributions as follows. A p-dimensioal random vector $\boldsymbol{Y}$ is Skew-normally distributed if $\boldsymbol{Y} = \boldsymbol{\mu} + U^{1/2}\boldsymbol{Z}$, where $\boldsymbol{\mu}$ is a location vector, U is a positive random variable with cumulative distribution function (cdf) $H(u|\boldsymbol{v})$ and probability density function (pdf) $h(u|\boldsymbol{v})$, and independent of the random vector $\boldsymbol{Z}$, $\boldsymbol{v}$ is a scalar or vector of parameters indexing the distribution of U, which is a positive value and $\boldsymbol{Z}$ is a multivariate Skew-normal

random vector with location $\mathbf{0}$, scale matrix $\mathbf{\Sigma}$ and Skewness parameter vector $\boldsymbol{\lambda}$, i.e. $\boldsymbol{Z} \sim SN_p(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda})$. When $U = u$, $\boldsymbol{Y}$ follows a multivariate Skew-normal distribution with location vector $\boldsymbol{\mu}$, scale matrix $u^{-1}\mathbf{\Sigma}$ and Skewness parameter vector $\boldsymbol{\lambda}$, i.e. $\boldsymbol{Y}|U = u \sim SN_p(\boldsymbol{\mu}, u^{-1}\mathbf{\Sigma}, \boldsymbol{\lambda})$. The marginal pdf of Y in that case is

$$f(\boldsymbol{y}) = 2\int_0^\infty \phi_p(\boldsymbol{y}; \boldsymbol{\mu}, u^{-1}\mathbf{\Sigma})\Phi(u^{-1/2}\boldsymbol{\lambda}^T\mathbf{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}))dH(u|\boldsymbol{v}) \qquad (1)$$

where $\phi_p(.; \boldsymbol{\mu}, \mathbf{\Sigma})$ denotes the pdf of the p-variate normal distribution $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and and $\Phi(.)$ represents the cdf of the standard univariate normal distribution. We will use the notation $\boldsymbol{Y} \sim SN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, H)$.

When $\lambda = 0$, the class of SN distributions reduces to the class of normal independent (N) distributions (Lachos et al. [15]; Lange and Sinsheimer [16]; Rosa et al. [27]), i.e., the class of thick-tailed distributions represented by the pdf

$$f_0(\boldsymbol{y}) = \int_0^\infty \phi_p(\boldsymbol{y}; \boldsymbol{\mu}, u^{-1}\mathbf{\Sigma})dH(u|\boldsymbol{v})$$

We will use the notation $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma}, H)$ for this case.

In the mixture model (1), when $U = 1$, $\boldsymbol{Y}$ is a multivariate Skew-normal distribution (SN) with location vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and and Skewness parameter vector $\boldsymbol{\lambda}$, i.e., $\boldsymbol{Y} \sim SN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda})$. The pdf of $\boldsymbol{Y}$ is

$$f(\boldsymbol{y}) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\mu}, \mathbf{\Sigma})\Phi(\boldsymbol{\lambda}^T\boldsymbol{y_0})$$

where $\boldsymbol{y_0} = \mathbf{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$.

A convenient stochastic representation of $\boldsymbol{Y}$ for simulation purposes, particularly data generation, follows from Bandyopadhyay et al. [3]-[4]:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta T} + \boldsymbol{\Gamma}^{-1/2}\boldsymbol{T_1} \qquad (2)$$

where $\boldsymbol{\Delta} = \mathbf{\Sigma}^{-1/2}\boldsymbol{\delta}$, $\boldsymbol{\Gamma} = \mathbf{\Sigma}^{1/2}(\boldsymbol{I} - \boldsymbol{\delta\delta}^T)\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma} - \boldsymbol{\Delta\Delta}^T$, $\boldsymbol{I}$ denotes the identity matrix and $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^T\boldsymbol{\lambda})^{1/2}$, $\boldsymbol{\lambda} = (\frac{\boldsymbol{\Gamma} + \boldsymbol{\Delta\Delta}^T)^{1/2}\boldsymbol{\Delta}}{[1 - \boldsymbol{\Delta}^T(\boldsymbol{\Gamma} + \boldsymbol{\Delta\Delta}^T)^1\boldsymbol{\Delta}]}$, $\mathbf{\Sigma} = \boldsymbol{\Gamma} + \boldsymbol{\Delta\Delta}^T$, $\boldsymbol{T} = |\boldsymbol{T_0}|$, $\boldsymbol{T_0} \sim N_1(0, 1)$ and $\boldsymbol{T_1} \sim N_p(\mathbf{0}, \boldsymbol{I_p})$.

When $\boldsymbol{U} \sim Gamma(\frac{v}{2}, \frac{v}{2})$, $v > 0$, $\boldsymbol{Y}$ follows a multivariate Skew-t distribution (St) with $\boldsymbol{v}$ degrees of freedom, i.e., $\boldsymbol{Y} \sim St_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{v})$. The pdf of $\boldsymbol{Y}$ is

$$f(\boldsymbol{y}) = 2t_p(\boldsymbol{y}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{v})T(\sqrt{\frac{v+p}{v+d}}\boldsymbol{A}; v+p), \boldsymbol{y} \in \mathbb{R}^p,$$

where $t_p(.; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{v})$ and $T(.; \boldsymbol{v})$ denote the pdf of the p-variate Student-t distribution and the cdf of the standard univariate t-distribution, respectively, $\boldsymbol{A} = \boldsymbol{\lambda}^T\mathbf{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$ and $\boldsymbol{d} = (\boldsymbol{y} - \boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$ is the Mahalanobis distance. As $\boldsymbol{v} \uparrow \infty$, we get the Skew-normal distribution. When $\boldsymbol{\lambda} = 0$, the Skew-t distribution reduces to the Student-t distribution.

2.2. The Linear Quadratic and nonlinear Spillman-Mitscherlich Model.

We first consider the general mixed model in which the random errors and the random parameters are independent and jointly normally distributed. The model reads:

$$Y_i = \eta_i(X_i, \varphi_i) + \epsilon, \varphi_i = A_i\beta + B_ib_i \tag{3}$$

where

$$b_i \sim N_q(0, Diag(D), \lambda, H), \epsilon_i \sim N_n(0, \sigma_e^2 I_{n_i}, H)$$

where the subscript i is the subject index, $i = 1, ..., n$; $Y_i = (y_{i1}, ..., y_{in_i})^T$ is a $n_i \times 1$ vector of observed continuous responses for sample unit i, $\eta_i(X_i, \varphi_i) = (\eta(X_{i1}, \varphi_i), ..., \eta(X_{n_i}, \varphi_i))^T$ with $\eta(.)$ the nonlinear or linear function of random parameters φ_i, and covariate vector X_i, A_i and B_i are known design matrices of dimensions $n_i \times p$ and $n_i \times q$, respectively, β is the $p \times 1$ vector of fixed parameter components (means), $b_i = (b_{1i}, ..., b_{qi})^T$ is the vector of random parameter components, and $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{in})^T$ is the vector of random errors, I_{n_i} denotes the identity matrix. The matrices $D = D(\alpha)$ with unknown parameter α is the $q \times q$ unstructured dispersion matrix of b_i, σ_e^2 the unknown variance of the error term and λ is the skewness parameter vector corresponding to the random components b_i.

We assume that $E(b_i) = E(\epsilon_i) = 0$, and the b_i and ϵ_i are uncorrelated, i.e. $Cov(b_i, \epsilon_i) = 0$. The model takes the within-subjects errors ϵ_i to be symmetrically distributed and the random parameter b_i to be asymmetrically distributed with mean zero (Bandyopadhyay et al. [4]; Lachos et al. [14]-[15]). When $\eta(.)$ is a nonlinear parametric function, we have the SN-NonLinear Mixed Model (SN-NLMM); if $\eta(.)$ is a linear parametric function, we have the SN-Linear Mixed Model (SN-LMM).

The general framework (3) gives the linear Quadratic and the nonlinear Spillman-Mitscherlich mixed model as follows:

1. The linear Quadratic mixed model:

$$Y_i = \gamma_1 + (\gamma_2 + b_{2i})X_i + (\gamma_3 + b_{3i})X_i^2 + b_{1i} + \epsilon_i, \tag{4}$$

where for $i = 1, 2, .., n$, Y_i is the response, X_i the dose, γ_1 is the intercept ; γ_2 the fixed linear response coefficient; γ_3 the fixed quadratic response coefficient; b_{1i}, b_{2i}, and b_{3i} are the random response coefficients; and ϵ_i is the random error term (Park et al. [22]; Tumusiime et. al. [34]). In this case, $\Gamma = (\gamma_1, \gamma_2, \gamma_3)^T$; $b_i = (b_{1i}, b_{2i}, b_{3i})^T$; $b_i \sim St_q(0, \Sigma, \lambda, v)$ and $\epsilon_i \sim t_q(0, \Sigma, v)$.

2. The nonlinear Spillman-Mitscherlich mixed model:

$$Y_i = \beta_1 + (\beta_2 + b_{2i})exp((-\beta_3 + b_{3i})X_i) + b_{1i} + \epsilon_i, \tag{5}$$

where the variables are as in (4), β_1 is the fixed maximum or potential response obtainable by the stimulus; β_2 is the fixed response increase induced by the stimulus; β_3 the ratio of successive increment in ouput β_1 to total output Y; b_{1i}, b_{2i}, and b_{3i} are the random components; and ϵ_i is the random error term (Tumusiime et. al. [34]); $\beta = (\beta_1, \beta_2, \beta_3)^T$; $b_i = (b_{1i}, b_{2i}, b_{3i})^T$; $b_i \sim St_q(0, \Sigma, \lambda, v)$ and $\epsilon_i \sim t_q(0, \Sigma, v)$.

3. BAYESIAN INFERENCE, GIBBS SAMPLER, AND SIMULATION EVALUATION
CRITERIA

3.1. **Prior distributions and joint posterior density.** As explained in the Introduction, we apply Bayesian inference to overcome the limitations of maximum likelihood. In spite of its advantages, Bayesian analysis also has some limitations. A mayor limitation which has hampered widespread implementation of the Bayesian approach, is that obtaining the posterior distribution often requires the integration of high-dimensional functions which can be analytically difficult. For simple models, the likelihood functions are standard, and, if one uses conjugate priors, deriving the posterior density analytically poses no major problems (This is the main reasons why conjugate priors are widely employed in Bayesian analysis). But Bayesian estimation quickly becomes challenging when working with more complicated models (possibly high-dimensional), or when one uses non-conjugate priors. Then analytical solution is not easy or may even be impossible. As a way out, Bayesian estimation using Markov Chain Monte Carlo (MCMC) simulation can be applied. Given a complex multivariate distribution, it is simpler to sample from a conditional distribution than to marginalize by integrating over a joint distribution. The MCMC approach proceeds on the basis of sampling from the complex distribution of interest. The algorithm departs from the previous sample value to generate the next sample value, thus generating a Markov chain. Specifically, let θ be the parameter of interest and let $y_1, ..., y_n$ be the numerical values of a sample from the distribution $f(y_1, ..., y_n|\theta)$. Suppose we sample (with replacement) some S independent, random θ-values from the posterior distribution $f(\theta|y_1, ..., y_n) : \theta^{(1)}, ..., \theta^{(S)} \sim$ i.i.d $f(\theta|y_1, ..., y_n)$. Then the empirical distribution of $\theta^{(1)}, ..., \theta^{(S)}$ approximates $f(\theta|y_1, ..., y_n)$ with the approximation improving with increasing S. The empirical distribution of $\theta^{(1)}, ..., \theta^{(S)}$ is known as a Monte Carlo approximation to $f(\theta|y_1, ..., y_n)$. Let $g(\theta)$ be (just about) any function of θ. The law of large numbers says that if $\theta^{(1)}, ..., \theta^{(S)}$ are i.i.d samples from $f(\theta|y_1, ..., y_n)$ then:

$$\frac{1}{S}\sum\nolimits_{s=1}^{S} g(\theta^s) \rightarrow E[g(\theta)|y_1, ..., y_n] = \int_{-\infty}^{\infty} g(\theta)f(\theta|y_1, ..., y_n)d\theta \qquad (6)$$

as $S \rightarrow \infty$. For further details we refer to Gelman et al. [8].

To further illustrate the procedure, consider a multidimensional parameter $\boldsymbol{\theta}$ for the case of n i.i.d observations from a Normal (μ, σ^2) with unknown mean and variance (i.e. $\boldsymbol{\theta} = (\mu, \sigma^2)$). We illustrate how to use the Markov Chain principle to simulate values from the joint posterior distribution $(\mu, \sigma^2|\boldsymbol{y})$, defined as the product of a conditional and a marginal distribution. This method is called decomposition method (Tanner [32]). Consider the density $f(\mu|\sigma^2, \boldsymbol{y})$. To obtain a sample $\mu^{(1)}, \mu^{(2)}..., \mu^{(S)}$ iid $f(\mu|y) = \int_{\sigma} f(\mu|\sigma^2, \boldsymbol{y})f(\sigma^2|\boldsymbol{y}) \, d\sigma^2$, we apply the composition method as follows:

    1. draw $\sigma^{2(s)}$ from $f(\sigma^2|\boldsymbol{y})$
    2. draw μ from $f(\mu|\sigma^{2(s)}, \boldsymbol{y})$

Steps 1 and 2 are repeated S times. The pairs $(\mu^{(1)}, \sigma^{2(1)}), .., (\mu^{(S)}, \sigma^{2(S)})$ are i.i.d samples from the joint density $f(\mu, \sigma^2|\boldsymbol{y}) = f(\mu|\sigma^2, \boldsymbol{y})f(\sigma^2|\boldsymbol{y})$, while the quantities $\mu^{(1)}, \mu^{(2)}..., \mu^{(S)}$ are i.i.d samples from the marginal $f(\mu|\boldsymbol{y})$.

For many multi parameter models the joint posterior distribution is non-standard (i.e. not a density like the normal or gamma distribution) and thus difficult to directly sample from. The composition method is impossible or hard to apply in that case because it is hard to get the marginal distributions of the random variables of interest. That is, it may be difficult to apply the decomposition method to generate independent observations from such a density $p(\theta|\boldsymbol{y})$, because the joint posterior distribution cannot be defined as the product of marginal and conditional distributions. An alternative solution in that case consists of generating a sample of correlated values which approximately come from the joint posterior distribution. Even if the sample observations are dependent, Monte Carlo integration can be applied, if the observations can be generated so that their joint density is roughly the same as the joint density of a random sample. The standard MCMC algorithms are:

- Metropolis
- Metropolis-Hasting
- Gibbs sampler

The Metropolis sampler obtains the state of the chain at $t + 1$ by sampling a candidate point θ_{new} from a proposal distribution $q(.|\theta^{(t)})$ which is only dependent on the previous state $\boldsymbol{\theta}^{(t)}$. The Metropolis algorithm can draw samples from any probability distribution $f(\theta|\boldsymbol{y})$ (target distribution), provided we can compute the value of a function $q(\theta|\boldsymbol{y})$ (proposal) that is proportional to the density of f. The lax requirement that $q(\theta|\boldsymbol{y})$ should be merely proportional to the target density, rather than exactly equal, makes the Metropolis algorithm particularly useful, because calculating the necessary normalization factor is often extremely difficult in practice. The algorithm works by generating a sequence of sample values in such a way that as more and more sample values are produced, the distribution of values more closely approximates the target distribution, $f(\theta|\boldsymbol{y})$. These sample values are produced iteratively with the distribution of the next sample being dependent only on the current sample value (thus making the sequence of samples a Markov chain). Specifically, at each iteration, the algorithm picks a candidate for the next sample value based on the current sample value. Then, with some probability, the candidate is either accepted (in which case the candidate value is used in the next iteration) or rejected (in which case the candidate value is discarded, and the current value is reused in the next iteration). The probability of acceptance is determined by comparing the current and candidate sample values of the function $q(\theta|\boldsymbol{y})$ with corresponding values of the target distribution $f(\theta|\boldsymbol{y})$.

The Metropolis sampler is based on a symmetric random-walk proposal distribution. A more general sampler is the Metropolis-Hastings algorithm which uses an asymmetric distribution: $q(\theta_{new}|\theta^{(t)}) \neq q(\theta^{(t)}|\theta_{new})$.

The Gibbs sampler is a special (simple) case of the Metropolis sampler in which the proposal distributions exactly match the posterior conditional distributions and proposals are accepted 100 % of the time. It decomposes the joint posterior distribution into full conditional distributions for each parameter in the model and then samples from them (A full conditional distribution is the conditional distribution of a parameter given all of the other parameters in the model).

The Gibbs sampler is efficient when the parameters are not highly dependent on each other and the full conditional distributions are easy to sample from. It is a popular sampling algorithm because it does not require a proposal distribution as the Metropolis method does, because the full conditional distribution is a standard distribution (e.g. normal or gamma). However, while deriving the full conditional distributions can be relatively easy, it is not always possible to find an efficient way to sample from these full conditional distributions. The Gibbs sampler proceeds as follows:

1. Set $t = 0$, and choose an arbitrary initial value of $\boldsymbol{\theta}^0 = \{\theta_1^{(0)}, \theta_2^{(0)}..., \theta_p^{(0)}\}$
2. Generate each vector component of $\boldsymbol{\theta}$ as follows:
    - draw $\theta_1^{(t+1)}$ from $f(\theta_1|\theta_2^{(t)}, ..., \theta_p^{(t)}, \boldsymbol{y})$
    - draw $\theta_2^{(t+1)}$ from $f(\theta_2|\theta_1^{(t)}, ..., \theta_p^{(t)}, \boldsymbol{y})$
    - draw ....
    - draw $\theta_p^{(t+1)}$ from $f(\theta_p|\theta_1^{(t)}, ..., \theta_{p-1}^{(t)}, \boldsymbol{y})$
3. Set t = t + 1. If t < T, i.e. the number of desired samples, return to step 2. Otherwise, stop.

Software such as JAGS (Just Another Gibbs Sampler) applies Gibbs sampling to implement Bayesian inference based on Markov Chain Monte Carlo simulation. In the Appendix A we present an example of an R program of the Gibbs sampler for a Bivariate distribution adapted from Rizzo [26].

The challenge of MCMC simulation is the construction of a Markov chain whose values converge to the target distribution. The general approach is the Metropolis-Hastings sampling procedure. This algorithm simulates samples from a probability distribution by making use of the full joint density function and independent proposal distributions for each of the variables of interest. Below, we apply the Gibbs sampler which is a special case of the Metropolis-Hastings sampling procedure. Gibbs sampling decomposes the joint posterior distribution into full conditional distributions for each parameter in the model and then samples from them. The proposal distributions in the Gibbs sampler exactly match the posterior conditional distributions. The sampler is usually efficient when the parameters are not highly dependent on each other and the full conditional distributions is easy to decompose.

Using (2), the mixed models under consideration can be formulated for $i = 1, .., n$, as follows (Lachos et al. [13]):

$$\boldsymbol{Y_i}|\boldsymbol{b_i}, U_i = u_i \sim N_{n_i}(\boldsymbol{\eta}(\boldsymbol{A_i\beta} + \boldsymbol{B_ib_i}, \boldsymbol{X_i}), u_i^{-1}\boldsymbol{\sigma}_e^2 I_{n_i})$$

$$\boldsymbol{b_i}|T_i = t_i, U_i = u_i \sim N_q(\boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma})$$

$$T_i|U_i = u_i \sim HN_1(0, u_i^{-1})$$
$$U_i \sim H(u_i|v)$$

where $HN_1(0, \boldsymbol{\sigma}^2)$ is the half-$N_1(0, \boldsymbol{\sigma}^2)$ distribution, $\boldsymbol{\Delta} = \boldsymbol{D}^{1/2}\boldsymbol{\delta}$ and $\boldsymbol{\Gamma} = \boldsymbol{D} - \boldsymbol{\Delta}\boldsymbol{\Delta}^T$, with $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^T\boldsymbol{\lambda})^{1/2}$ and $\boldsymbol{D}^{1/2}$ the square root of $\boldsymbol{D}$ containing $q(q+1)/2$ distinct elements.

Let $\boldsymbol{Y} = (\boldsymbol{y_1}^T, ..., \boldsymbol{y_n}^T)^T$, $\boldsymbol{b_i} = (\boldsymbol{b_1}^T, ..., \boldsymbol{b_n}^T)^T$, $\boldsymbol{u} = (u_1, ..., u_n)^T$, $\boldsymbol{t} = (t_1, ..., t_n)^T$. Then, the complete likelihood function associated with $(\boldsymbol{y}^T, \boldsymbol{b}^T, \boldsymbol{u}^T, \boldsymbol{t}^T)^T$ is given by

$$L(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{u}, \boldsymbol{t}) \propto \prod_{i=1}^{n}[\phi_{\boldsymbol{n_i}}(\boldsymbol{y_i}; \boldsymbol{\eta}(\boldsymbol{A_i}\boldsymbol{\beta} + \boldsymbol{B_i}\boldsymbol{b_i}, \boldsymbol{X_i}), u_i^{-1}\boldsymbol{\sigma}_e^2 I_{n_i})\phi_{\boldsymbol{q}}(\boldsymbol{b_i}; \boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma})$$

$$\times \phi_{\boldsymbol{1}}(t_i; 0, u_i^{-1})h(u_i|v)]$$

The unknown parameters of this model are $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\sigma}_e^2, \boldsymbol{\alpha}^T, \boldsymbol{\lambda}^T, \boldsymbol{v}^T)^T$.

In the Bayesian framework, we need to consider prior distributions for all the unknown parameters. We consider $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta_0}, \boldsymbol{S_\beta})$, $\boldsymbol{\sigma}_e^2 \sim IG(\boldsymbol{\omega_1}, \boldsymbol{\omega_2})$, $\boldsymbol{\Gamma} \sim IW_q(\boldsymbol{M_0}, l)$, $\boldsymbol{\Delta} \sim N_p(\boldsymbol{\Delta_0}, \boldsymbol{S_\Delta})$, where $IG(\boldsymbol{\omega_1}, \boldsymbol{\omega_2})$ is the inverse gamma distribution with mean $\boldsymbol{\omega_2}((\boldsymbol{\omega_1} - 1))$, $\boldsymbol{\omega_1} > 1$, and $IW_q(\boldsymbol{M}, l)$ is the inverse Wishart distribution with mean $\boldsymbol{M}/(l-q-1))$, $l > q+1$, where $\boldsymbol{M}$ is a $q \times q$ known positive definite matrix (Bandyopadhyay et al. [3]; Lachos et al. [13]). Then, the joint prior distribution of all unknown parameters is

$$\boldsymbol{\pi}(\boldsymbol{\theta}) = \boldsymbol{\pi}(\boldsymbol{\beta})\boldsymbol{\pi}(\boldsymbol{\sigma}_e^2)\boldsymbol{\pi}(\boldsymbol{\Gamma})\boldsymbol{\pi}(\boldsymbol{\Delta})\boldsymbol{\pi}(\boldsymbol{v})$$

Combining the likelihood function and the prior distribution, the joint posterior density of all unknown is

$$\boldsymbol{\pi}(\boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Gamma}, \boldsymbol{\Delta}, \boldsymbol{b}, \boldsymbol{u}, \boldsymbol{t}|\boldsymbol{y}) \propto \prod_{i=1}^{n}[\phi_{\boldsymbol{n_i}}(\boldsymbol{y_i}; \boldsymbol{\eta}(\boldsymbol{A_i}\boldsymbol{\beta} + \boldsymbol{B_i}\boldsymbol{b_i}, \boldsymbol{X_i}), u_i^{-1}\boldsymbol{\sigma}_e^2 I_{n_i})$$

$$\phi_{\boldsymbol{q}}(\boldsymbol{b_i}; \boldsymbol{\Delta}t_i, u_i^{-1}\boldsymbol{\Gamma}) \times \phi_{\boldsymbol{1}}(t_i; 0, u_i^{-1})h(u_i|v)]\boldsymbol{\pi}(\boldsymbol{\theta})$$

Under this full model, given $\mathbf{u}$, the full conditional distribution of $\boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{b_i}$, $t_i, i = 1, .., n$, are as follows (Lachos et al. [13]):

$\boldsymbol{\beta}|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Delta}, \boldsymbol{\Gamma} \sim N_p(\boldsymbol{\mathcal{A}}_\beta^{-1}\mathbf{a}_\beta, \boldsymbol{\mathcal{A}}_\beta^{-1})$,
where $\boldsymbol{\mathcal{A}}_\beta = \boldsymbol{\mathcal{S}}_\beta^{-1} + (\frac{1}{\sigma_e^2})\sum_{i=1}^{n}u_i^{-1}\mathbf{X}_i^T\mathbf{I}_n^{-1}\mathbf{X}_i$ and $\mathbf{a}_\beta = \boldsymbol{\mathcal{S}}_\beta^{-1}\boldsymbol{\beta_0} + (\frac{1}{\sigma_e^2})\sum_{i=1}^{n}u_i^{-1}\mathbf{X}_i^T\mathbf{I}_n^{-1}$ $(\boldsymbol{y_i} - \boldsymbol{B_i}\boldsymbol{b_i})$

$\boldsymbol{\sigma}_e^2|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma} \sim IG(\frac{N+\tau_e}{2}, \frac{T_e\sum_{i=1}^{n}u_i^{-1}\boldsymbol{\mu}_i^T\mathbf{I}_n^{-1}\boldsymbol{\mu}_i}{2})$ where $N = \sum_{i=1}^{n}n_i$ and $\boldsymbol{\mu}_i = \boldsymbol{y_i} - \boldsymbol{A_i}\boldsymbol{\beta} - \boldsymbol{B_i}\boldsymbol{b_i}$

$\boldsymbol{\Delta}|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Gamma} \sim N(\sum_{\Delta}^{-1}\boldsymbol{\mu}_\Delta, \sum_{\Delta}^{-1})$,
where $\boldsymbol{\mu}_\Delta = \boldsymbol{S}_\Delta^{-1}\boldsymbol{\Delta_0} + \boldsymbol{\Gamma}^{-1}\sum_{i=1}^{n}u_i\boldsymbol{b_i}$, $\sum_\Delta = \boldsymbol{\Gamma}^{-1}\sum_{i=1}^{n}u_i\boldsymbol{t}_i^2 + \boldsymbol{S}_\Delta^{-1}$

$\boldsymbol{\Gamma}|\mathbf{b}, \mathbf{u}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Delta} \sim IW_{\tau_b+n}((T_b^{-1} + \sum_{i=1}^{n}u_i(\boldsymbol{b_i} - \boldsymbol{\Delta}t_i)(\boldsymbol{b_i} - \boldsymbol{\Delta}t_i)^T)^{-1})$

$\boldsymbol{b_i}|\boldsymbol{\beta}, \boldsymbol{\sigma}_e^2, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \mathbf{u}_i, \boldsymbol{t}_i \sim N_q(\boldsymbol{\mathcal{A}}_{b_i}^{-1}\mathbf{a}_i, u_i^{-1}\boldsymbol{\mathcal{A}}_{b_i}^{-1})$,
where $\boldsymbol{\mathcal{A}}_{b_i} = ((\frac{1}{\sigma_e^2})\mathbf{B}_i^T\mathbf{I}_n^{-1}\mathbf{B}_i + \boldsymbol{\Gamma}^{-1})$ and $\mathbf{a}_i = (\frac{1}{\sigma_e^2})\mathbf{B}_i^T\mathbf{I}_n^{-1}(\boldsymbol{y_i} - \boldsymbol{A_i}\boldsymbol{\beta}) + t_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}, i = 1, .., n,$

$$\boldsymbol{T}_i|\boldsymbol{\beta},\boldsymbol{\sigma}_e^2,\boldsymbol{\Gamma},\boldsymbol{\Delta},\boldsymbol{b}_i,u_i \sim N(\boldsymbol{\mathcal{A}}_t^{-1}\boldsymbol{\mathfrak{a}}_{t_i},\mathbf{u}_i^{-1}\boldsymbol{\mathcal{A}}_t^{-1})\|\{\boldsymbol{T}_i > 0\},$$

where $\boldsymbol{\mathcal{A}}_t = (1+\boldsymbol{\Delta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})$ and $\boldsymbol{\mathfrak{a}}_{ti} = (1+\boldsymbol{b_i}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})$, $i=1,..,n$. Then,

$$\boldsymbol{D} = \boldsymbol{\Gamma}+\boldsymbol{\Delta}\boldsymbol{\Delta}^T \text{ and } \boldsymbol{\lambda} = \boldsymbol{D}^{-1/2}\boldsymbol{\Delta}/(1-\boldsymbol{\Delta}^T\boldsymbol{D}^{-1}\boldsymbol{\Delta})^{1/2}$$

For Skew-t,

$$u_i|\boldsymbol{\theta},\boldsymbol{y},\boldsymbol{b},\boldsymbol{t} \sim Gamma(\tfrac{(n_i+q+v+1)}{2};\tfrac{v}{2}+\tfrac{C_i}{2}),$$

where $C_i = \frac{1}{\sigma_e^2}(\boldsymbol{y_i}-\boldsymbol{A_i}\boldsymbol{\beta}-\boldsymbol{B_i}\boldsymbol{b_i})^T\boldsymbol{R}_i^{-1}(\boldsymbol{y_i}-\boldsymbol{A_i}\boldsymbol{\beta}-\boldsymbol{B_i}\boldsymbol{b_i})+(\boldsymbol{b_i}-\boldsymbol{\Delta}t_i)^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{b_i}-\boldsymbol{\Delta}t_i)+\boldsymbol{t}_i^2$, and

$$\boldsymbol{\pi}(v|\boldsymbol{\theta}_{-v},\boldsymbol{y},\boldsymbol{b},\boldsymbol{u},\boldsymbol{t}) \propto (2^{\frac{v}{2}}\boldsymbol{\Gamma}(\tfrac{v}{2}))^{-n}v^{\frac{nv}{2}}exp(-\tfrac{v}{2}[\textstyle\sum_{i=1}^n(u_i-logu_i)+\varrho])\,\|_{2,\boldsymbol{\infty}}.$$

3.2. **Model comparison criteria.** For model comparison, we use the deviance information criterion (DIC), the expected Akaike information criterion (EAIC) and the expected Bayesian information criterion (EBIC). These are based on the posterior mean of the deviance which can be approximated as $\bar{\mathcal{D}} = \sum_{q=1}^Q \mathcal{D}(\boldsymbol{\theta_q})/Q$, where $\mathcal{D}(\boldsymbol{\theta}) = -2\sum_{i=1}^n logf(y_i|\boldsymbol{\theta})$ and Q is the number of iterations. The EAIC, EBIC and DIC can be estimated using MCMC output as follows

$$\widehat{EAIC} = \bar{\mathcal{D}}+2p, \widehat{EBIC} = \bar{\mathcal{D}}+plog(N), \widehat{DIC} = \bar{\mathcal{D}}+p_v$$

where $\bar{\mathcal{D}}$ is the posterior mean of the deviance, p the number of parameters in the model, N the total number of observations, and p_v the effective number of parameters defined as *Variance ($\mathcal{D}$)/2* (Plummer [25]; Spiegelhalter et al. [29]-[30]).

4. SIMULATION SETUP

In the simulations, we consider the two most common dose-response models, i.e., the linear Quadratic model and the nonlinear Spillman-Mitscherlich model (Models (4) and (5), respectively).

We generate 100 Monte Carlo simulation datasets (samples). The number of 100 samples is chosen because of the long processing time of Bayesian estimation. We consider three sample sizes: $(T) =$10, 30, 75, i.e. a small, medium and a large number of observations, respectively. In each group, the following six doses are applied: 0, 100, 150, 200, 300, 400. To avoid non-convergence, we normalize the original doses (subtract the mean and divide by the standard deviation) (Kery [12]). The simulation were performed using the following fixed parameter values $\boldsymbol{\beta}^T = (\beta_1,\beta_2,\beta_3)$=(8.0,1.5,0.5), $\boldsymbol{\gamma}^T = (\gamma_1,\gamma_2,\gamma_3)$=(6.0,0.5,0.25).

The following A_i and B_i matrices were applied

$$\mathbf{A_i} = \mathbf{B_i} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The scale matrix of the random components is

$$\mathbf{D} = \begin{bmatrix} \sigma_{b_1} & 0 & 0 \\ 0 & \sigma_{b_2} & 0 \\ 0 & 0 & \sigma_{b_3} \end{bmatrix}$$

To get insight into the performance of the estimators under increasing variance, we analyzed small, medium and large scale $\mathbf{D}$ matrices (scenario 1-3) as follows:

Simulation	σ_{b_1}	σ_{b_2}	σ_{b_3}	σ_e
Scenario 1	0.1	0.01	0.005	0.5
Scenario 2	1	0.1	0.05	1
Scenario 3	1.5	0.2	0.10	0.75

To analyze the Skew-t distributions, we generated $\beta_k + b_{ki}$ and $\gamma_k + b_{ki}$, $k = 1, 2, 3$ according to the multivariate (right) Skew-t distribution $St_3 \sim (0, \sigma_{b_k}, 3, 4)$ and the ϵ_i according to the t-distribution $\epsilon_i \sim t_1(0, \sigma^2, 4)$. For the multivariate normal distributions, we generated $\beta_k + b_{ki}$ and $\gamma_k + b_{ki}$ according to the multivariate normal distribution $N_3 \sim (0, \sigma_{b_k})$ and the ϵ_i according to the normal distribution $\epsilon_i \sim N_1(0, \sigma^2)$.

For each of the 100 simulated data sets, the linear Quadratic and the Spillman-Mitscherlich random parameter models were estimated under the assumption that (1) the density of random components was the Skew-t and the density of the errors the t distribution, (2) the random components and the errors were normally distributed (N).

The following independent priors were considered to analyze the Gibbs sampler : $\beta_k \sim N(0, 10^3)$, $\gamma_k \sim N(0, 10^3)$, $\sigma_e{}^2 \sim IG(0.01, 0.01)$, $\mathbf{\Gamma} \sim IW_3(H)$ with $H = diag(0.01)$ for the normal, Skew-normal, and Skew-t priors, and $\mathbf{\Delta} \sim N(0, 0.001)$ for the Skew-normal and Skew-t priors and $v \sim Exp(0.1; (2, \infty))$ for the Skew-t prior. For these prior densities, we generated two parallel independent runs of the Gibbs sampler chain of size 25 000 for each parameter. We disregarded the first 5 000 iterations to eliminate the effect of the initial value. To avoid potential auto-correlation, we used a thinning of 10. We assessed chain convergence using trace plots, autocorrelation plots and the Brooks-Gelman-Rubin scale reduction factor $\hat{R}$ (Gelman et al. [8]). We fitted the models using the R2jags package available in R (Su and Yajima [31]). We computed the relative bias ($RelBias$) and the Root Mean Square Error ($RMSE$) for each parameter estimate over 100 samples for each simulation. These statistics are defined as

$$RelBias(\boldsymbol{\theta}) = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\hat{\boldsymbol{\theta}_j}}{\boldsymbol{\theta}} - 1\right), RMSE(\boldsymbol{\theta}) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{\boldsymbol{\theta}_j} - \boldsymbol{\theta})^2}$$

where $\hat{\boldsymbol{\theta}_j}$ is the estimate of $\boldsymbol{\theta}$ for the jth sample and N=100.

5. Main Results

5.1. **(Right) Skewed-*t* response data.** Tables 1 and 2 and Figures 1 show that for the nonlinear Spillman-Mitscherlich model and right-skewed, heavy-tailed response data, the average *RelBias* (in absolute value) and the average *RMSE* for all T, **D**, and the three parameters of the normal prior (N) are larger than for the Skew normal prior (SN). However, for the linear Quadratic model the opposite holds. Moreover, the average *RelBias* (in absolute value) and the average *RMSE* of the Skew-*t* priors (S-*t*) have the smallest values for all sample sizes (*T*) and for the three variance for both models.

FIGURE 1. Average *RelBias* and *RMSE* of the Normal (N), Skew Normal (SN), and Skew-*t* (S-*t*) prior for the Spillman-Mitscherlich and linear Quadratic model for right-skewed data

Note also that are for the Spillman-Mitscherlich model with T=10 and Small-**D** the *RelBias* of β_3 in the case of the normal prior is smaller than in the case of the Skew-normal and Skew-*t* priors, while for large-**D** the *RelBias* of β_1 and β_2 for the three priors are similar. Moreover, for the Spillman-Mitscherlich model for large-**D** and an increasing number of observations up to 30 the *RelBias* and *RMSE* of β_1 and β_2 of the three priors worsen, but improve for increasing T. This sample size bias inconsistency was also observed by Hagenbuch [9]. Apparently, one should increase the sample size substantially to reduce the bias.

From the above it follows that except for some minor exceptions, for skewed heavy tailed data the Bayesian estimator with Skew-*t* prior is more accurate than in the case of the normal and the Skew normal prior. (Note that because of different scales, inferences regarding the variance component are not feasible (Lachos et al. [13]-[14]).
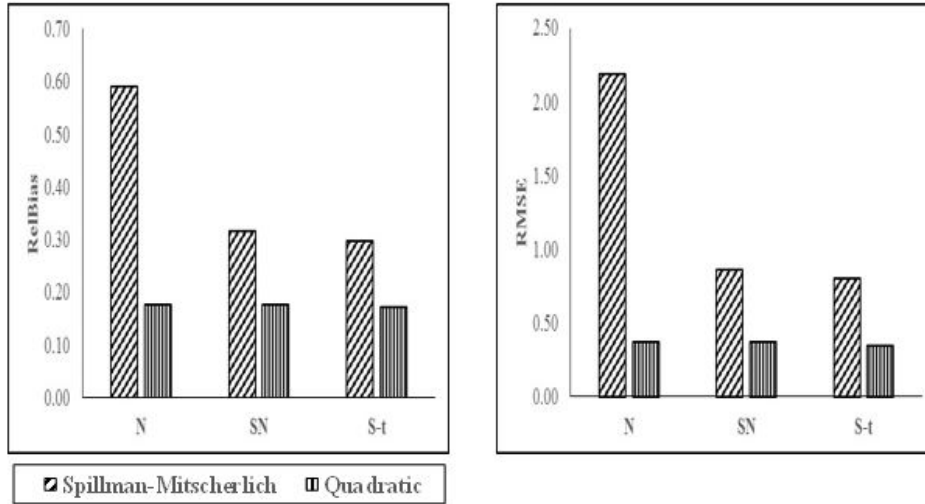
TABLE 1. *RelBias* and *RMSE* of the Normal (N), Skew-Normal (SN), and Skew-t (S-t) prior for the Spillman-Mitscherlich model for right-skewed data

(T)	Variance (**D**)	Parameter	Prior					
			N		SN		S-t	
			RelBias	RMSE	RelBias	RMSE	RelBias	RMSE
10	Small	β_1	0.1176	1.8756	0.0002	0.4154	0.0026	0.3940
		β_2	0.6150	1.8844	-0.0358	0.4275	-0.0259	0.4249
		β_3	-0.0734	0.1821	0.1441	0.1640	0.1475	0.1620
	Medium	β_1	-0.0269	2.7153	-0.0555	1.1592	-0.056	1.2466
		β_2	-0.2866	2.6967	-0.4600	1.1752	-0.4653	1.2662
		β_3	-0.5596	0.4212	-0.3843	0.4089	-0.3734	0.4147
	Large	β_1	-0.0192	2.0702	0.0116	1.1706	-0.0049	1.0671
		β_2	-1.2126	2.7233	-1.0534	1.9091	-1.0858	1.9133
		β_3	-0.8753	0.4724	-0.8314	0.4709	-0.9243	0.5039
30	Small	β_1	0.1161	1.6831	0.0307	0.4693	0.0313	0.4328
		β_2	0.584	1.6802	0.1093	0.4514	0.1186	0.4073
		β_3	-0.1684	0.1555	-0.0061	0.1087	-0.0317	0.0927
	Medium	β_1	-0.0325	2.6284	0.0364	0.3674	0.0232	0.2936
		β_2	-0.8435	2.8891	-0.4728	0.7367	-0.4561	0.7140
		β_3	1.1832	0.7182	0.8151	0.4346	0.7878	0.4177
	Large	β_1	1.2093	10.982	0.5926	4.9354	0.5144	4.3627
		β_2	5.9196	10.275	2.5624	4.0886	2.1467	3.5399
		β_3	-0.7918	0.4131	-0.4565	0.2483	-0.3838	0.2209
75	Small	β_1	0.0206	1.0500	0.0147	0.3906	0.0158	0.3129
		β_2	0.0455	1.0412	0.0055	0.3911	0.0131	0.3028
		β_3	-0.0453	0.1116	0.0246	0.0907	0.0047	0.0724
	Medium	β_1	0.1626	2.4197	0.0955	0.8902	0.101	0.9232
		β_2	0.3881	2.1373	0.0029	0.4884	0.0461	0.4772
		β_3	-0.1465	0.2157	0.0908	0.1412	0.0495	0.1285
	Large	β_1	0.1479	2.8697	0.1273	1.0880	0.1234	1.0609
		β_2	0.0988	2.6152	-0.0452	0.4315	-0.0326	0.3910
		β_3	-0.2397	0.2417	0.0659	0.1290	0.0255	0.1187
	Average	$\boldsymbol{\beta}$	0.5900	2.1916	0.3159	0.8586	0.2960	0.8023

Tables 3 and 4 show the overall fit statistics for the Spillman-Mitscherlich and linear Quadratic model. For both models the DIC, EAIC, and EBIC all tend to favor the Skew-t model for all sample sizes (T) and for the three variance scenarios. The percentage (%) of samples that the criteria choose the Skew-t model as the best model increases with increasing number of observations. For T=75 the percentage is 100%.

Table 2. *RelBias* and *RMSE* of the Normal (N), Skew-Normal (SN), and Skew-*t* (S-*t*) prior for the linear Quadratic model for right-skewed data

			Prior					
(T)	Variance (**D**)	Parameter	N		SN		S-*t*	
			RelBias	RMSE	RelBias	RMSE	RelBias	RMSE
10	Small	γ_1	0.0123	0.1610	0.0124	0.1613	0.0127	0.1379
		γ_2	0.0353	0.1002	0.0350	0.1011	0.0289	0.0869
		γ_3	0.0105	0.1104	0.0129	0.1123	0.0046	0.0945
	Medium	γ_1	0.2059	1.2626	0.2066	1.2671	0.1748	1.0791
		γ_2	0.1410	0.2357	0.1485	0.2399	0.1154	0.2152
		γ_3	-0.1903	0.2174	-0.1998	0.2189	-0.2698	0.2034
	Large	γ_1	0.3223	1.9423	0.3223	1.9425	0.3167	1.9127
		γ_2	0.3966	0.2333	0.3971	0.2355	0.3867	0.2313
		γ_3	-0.2864	0.1588	-0.2917	0.1629	-0.2654	0.1541
30	Small	γ_1	0.0181	0.1341	0.0182	0.135	0.0185	0.1325
		γ_2	0.0403	0.0602	0.0413	0.0611	0.0458	0.0523
		γ_3	-0.0114	0.0598	-0.0116	0.0602	0.0017	0.0514
	Medium	γ_1	0.0898	0.5616	0.0897	0.5609	0.0882	0.5492
		γ_2	0.2959	0.1927	0.2960	0.1944	0.2936	0.1807
		γ_3	-0.1460	0.1179	-0.1494	0.1216	-0.1353	0.1057
	Large	γ_1	0.2046	1.2333	0.2048	1.2346	0.1971	1.1905
		γ_2	0.4419	0.2331	0.4431	0.2338	0.4281	0.2260
		γ_3	-0.4560	0.1376	-0.4587	0.1386	-0.4719	0.1409
75	Small	γ_1	0.0166	0.1109	0.0167	0.1111	0.0155	0.1029
		γ_2	0.0272	0.0369	0.0273	0.0368	0.0248	0.0337
		γ_3	0.0159	0.0350	0.0164	0.0353	0.0087	0.0298
	Medium	γ_1	0.1006	0.6110	0.1007	0.6114	0.0941	0.5722
		γ_2	0.1723	0.1112	0.1726	0.1115	0.165	0.1052
		γ_3	-0.2113	0.0896	-0.2121	0.0900	-0.2137	0.0863
		γ_1	0.2488	1.4944	0.2487	1.4943	0.2395	1.4391
	Large	γ_2	0.2914	0.1533	0.2919	0.1539	0.2872	0.1506
		γ_3	-0.3444	0.1020	-0.3457	0.1029	-0.3573	0.1024
	Average	$\boldsymbol{\Gamma}$	0.1753	0.3665	0.1767	0.3677	0.1726	0.3469

Note also that for Spillman-Mitscherlich data, T=10 and Small-**D**, the DIC selects the S-*t* model as the best model while the EAIC and EBIC select the normal model. For T=30 and large-**D**, all the measures favor the normal model. The different results are probably a consequence of the fact that the three measures penalize model complexity differently. According to Spiegelhalter et al. [30]), the AIC is based on the number of parameters, the BIC on the log sample size and the DIC on the effective number of parameters. i.e. on $p_D = E\mathcal{D}(\boldsymbol{\theta}) - \mathcal{D}E(\boldsymbol{\theta})$,

TABLE 3. *DIC*, *EAIC* and *EBIC* for Normal (N), Skew Normal (SN), and Skew-*t* (S-*t*) priors for the Spillman-Mitscherlich model for right-skewed data

			Prior		
(T)	Variance (**D**)	Parameter	N	SN	S-*t*
10	Small	DIC	215.91 (33%)	218.30 (0%)	213.26 (67%)
		EAIC	222.91 (65%)	228.30 (1%)	224.26 (34%)
		EBIC	221.35 (59%)	226.08 (0%)	221.82 (41%)
	Medium	DIC	126.67 (24%)	128.85 (2%)	120.73 (74%)
		EAIC	133.67 (51%)	138.85 (0%)	131.73 (49%)
		EBIC	132.12 (47%)	136.64 (0%)	129.29 (53%)
	Large	DIC	251.92 (57%)	254.27 (1%)	251.87 (42%)
		EAIC	258.92 (84%)	264.27 (0%)	262.87 (16%)
		EBIC	257.37 (81%)	262.05 (0%)	260.43 (19%)
30	Small	DIC	688.41 (6%)	692.20 (1%)	667.41 (93%)
		EAIC	695.41 (15%)	702.20 (1%)	678.41 (94%)
		EBIC	697.19 (18%)	704.75 (1%)	681.22 (81%)
	Medium	DIC	374.30 (0%)	372.68 (0%)	339.74 (100%)
		EAIC	381.30 (0%)	382.68 (0%)	350.74 (100%)
		EBIC	383.08 (0%)	385.23 (0%)	353.55 (100%)
	Large	DIC	694.82 (59%)	702.94 (0%)	695.84 (41%)
		EAIC	701.82 (73%)	712.94 (0%)	706.84 (27%)
		EBIC	703.60 (79%)	715.50 (0%)	709.65 (21%)
75	Small	DIC	1696.27 (0%)	1692.25 (0%)	1635.75 (100%)
		EAIC	1703.27 (0%)	1702.25 (0%)	1646.75 (100%)
		EBIC	1707.84 (0%)	1708.78 (0%)	1653.93 (100%)
	Medium	DIC	943.94 (0%)	933.49 (0%)	849.52 (100%)
		EAIC	950.94 (0%)	943.49 (0%)	860.52 (100%)
		EBIC	955.52 (0%)	950.02 (0%)	867.71 (100%)
	Large	DIC	1746.52 (0%)	1742.90 (0%)	1686.90 (100%)
		EAIC	1753.52 (0%)	1752.90 (0%)	1697.90 (100%)
		EBIC	1758.09 (0%)	1759.43 (0%)	1705.08 (100%)

Note: Within brackets the percentage that the criterion selected the given prior

where $E\mathcal{D}(\boldsymbol{\theta})$ is the posterior mean of the deviance and $E\mathcal{D}(\boldsymbol{\theta})$ is the deviance of posterior mean of the model parameters. Some studies showed that compared to DIC, the AIC and BIC favor simpler models (i.e. with less parameters) (Ward [36]; Spiegelhalter et al. [30]).

The above results show that the average *RelBias* (in absolute value) and the average *RMSE* (for all T, **D**, and the three parameters) of the Spillman Mitscherlich model are larger than of the linear Quadratic model. Moreover, for both models

Table 4. *DIC*, *EAIC* and *EBIC* for the Normal (N), Skew
Normal (SN), and Skew-*t* (S-*t*) priors for the linear Quadratic
model for right-skewed data

			Prior		
(T)	Variance (**D**)	Parameter	N	SN	S-*t*
10	Small	DIC	213.16 (29%)	216.51 (0%)	206.97 (71%)
		EAIC	220.16 (52%)	226.51 (0%)	217.97 (48%)
		EBIC	218.61 (48%)	224.30 (0%)	215.53 (52%)
	Medium	DIC	127.14 (17%)	129.10 (1%)	119.31 (82%)
		EAIC	134.14 (44%)	139.10 (0%)	130.31 (56%)
		EBIC	132.58 (41%)	136.89 (0%)	127.87 (59%)
	Large	DIC	174.10 (22%)	176.52 (1%)	168.55 (77%)
		EAIC	181.10 (45%)	186.52 (0%)	179.55 (55%)
		EBIC	179.54 (39%)	184.30 (0%)	177.11 (61%)
30	Small	DIC	633.77 (0%)	637.40 (0%)	604.50 (100%)
		EAIC	640.77 (4%)	647.40 (0%)	615.50 (96%)
		EBIC	642.56 (4%)	649.95 (0%)	618.31 (96%)
	Medium	DIC	380.70 (0%)	379.21 (0%)	337.54 (100%)
		EAIC	387.70 (1%)	389.21 (0%)	348.54 (99%)
		EBIC	389.49 (1%)	391.76 (0%)	351.35 (99%)
	Large	DIC	530.48 (0%)	530.74 (1%)	501.59 (99%)
		EAIC	537.48 (3%)	540.74 (1%)	512.59 (96%)
		EBIC	539.27 (5%)	543.29 (1%)	515.39 (94%)
75	Small	DIC	1572.52 (0%)	1573.29 (0%)	1488.31 (100%)
		EAIC	1579.52 (0%)	1583.29 (0%)	1499.31 (100%)
		EBIC	1584.10 (0%)	1589.82 (0%)	1506.49 (100%)
	Medium	DIC	945.79 (0%)	937.11 (0%)	837.70 (100%)
		EAIC	952.79 (0%)	947.11 (0%)	848.70 (100%)
		EBIC	957.36 (0%)	953.64 (0%)	855.88 (100%)
	Large	DIC	1319.80 (0%)	1314.63 (0%)	1232.31 (100%)
		EAIC	1326.80 (0%)	1324.63 (0%)	1243.31 (100%)
		EBIC	1331.37 (0%)	1331.16 (0%)	1250.50 (100%)

Note: Within brackets the percentage that the corresponding criterion favored
the corresponding prior

the DIC, EAIC, and EBIC all tend to favor the Skew-*t* model for all sample sizes
(T) and for the three variance scenarios, although there are some minor exceptions
for the Spillman-Mitscherlich model. These results indicate a major drawback of
the nonlinear mixed model. According to Harring and Liu [10] estimation - in-
cluding Bayesian estimation - of model parameters of nonlinear mixed model are
not straightforward compared to its counterpart, the linear mixed model. The
nonlinearity requires multidimensional integration to derive the needed marginal

TABLE 5. *RelBias* and *RMSE* of the Normal (N), Skew-Normal (SN), and Skew-t (S-t) prior for the Spillman-Mitscherlich model for normal data

(T)	Variance (**D**)	Parameter	Prior					
			N		SN		S-t	
			RelBias	RMSE	RelBias	RMSE	RelBias	RMSE
10	Small	β_1	0.0524	2.0806	0.0013	0.4082	0.0021	0.4231
		β_2	0.3208	2.0967	0.0232	0.4293	0.0275	0.4477
		β_3	-0.1728	0.1620	0.0618	0.1145	0.0645	0.1249
	Medium	β_1	-0.3815	4.5593	-0.2038	2.0963	-0.2843	2.6841
		β_2	-1.8057	4.3644	-0.878	1.7884	-1.272	2.3859
		β_3	-0.8085	0.4653	-0.6716	0.4773	-0.8831	0.6009
	Large	β_1	0.0205	2.5966	-0.0168	0.6133	-0.0199	0.6950
		β_2	-0.0366	2.2582	-0.2755	0.8334	-0.2879	0.7314
		β_3	-0.2319	0.2327	0.0764	0.1944	0.0194	0.2644
30	Small	β_1	0.0158	0.8942	-0.0039	0.2975	-0.0023	0.3064
		β_2	0.0813	0.9019	-0.0337	0.3221	-0.0246	0.3297
		β_3	-0.0390	0.1134	0.0463	0.0930	0.0389	0.0942
	Medium	β_1	1.1092	9.5134	0.3338	2.7361	0.3333	2.7289
		β_2	5.8485	9.4216	1.6683	2.5723	1.6503	2.5427
		β_3	-0.7840	0.3965	-0.5016	0.2526	-0.4981	0.2508
	Large	β_1	-0.1458	1.2162	-0.1294	1.0389	-0.1446	1.1601
		β_2	-0.6243	0.9940	-0.5159	0.7779	-0.5956	0.8960
		β_3	1.2061	0.6270	0.7472	0.3821	0.8216	0.4208
75	Small	β_1	0.0006	0.2267	-0.0066	0.1879	-0.0072	0.1900
		β_2	0.0017	0.2442	-0.0403	0.2030	-0.0428	0.2052
		β_3	0.0227	0.0648	0.0471	0.0613	0.0489	0.0622
	Medium	β_1	0.0765	1.2537	-0.0094	0.3512	-0.0012	0.3577
		β_2	0.4821	1.3349	0.0059	0.3589	0.0434	0.3772
		β_3	-0.1185	0.1420	0.0321	0.0961	0.0122	0.0946
	Large	β_1	0.1835	2.3119	0.1385	1.1888	0.1387	1.1823
		β_2	0.8173	1.6597	0.5605	0.9529	0.5721	0.9577
		β_3	-0.2468	0.1447	-0.1475	0.1015	-0.1663	0.1117
	Average	$\boldsymbol{\beta}$	0.5791	1.8621	0.2658	0.7011	0.2964	0.7639

distribution of the data from which inferences can be made. The integral is almost always intractable, i.e. has no closed form solution.

5.2. **Normal data.** From Table 5 and Figures 2 it follows that in the case of the Spillman-Mitscherlich model and normally distributed data, the normal prior average *RelBias* (in absolute value) and the average *RMSE* of the over all T, **D**, and the three parameters are larger than of the Skew normal prior and Skew-t prior. Furthermore, the average *RelBias* (in absolute value) and average *RMSE*

TABLE 6. *RelBias* and *RMSE* for the Normal (N), Skew-Normal (SN), and Skew-*t* (S-*t*) prior for the linear Quadratic model for normal data

(T)	Variance (**D**)	Parameter	Prior					
			N		SN		S-*t*	
			RelBias	RMSE	RelBias	RMSE	RelBias	RMSE
10	Small	γ_1	0.0053	0.0977	0.0053	0.0977	0.0053	0.0978
		γ_2	0.0021	0.0663	0.0017	0.0663	0.0010	0.0706
		γ_3	0.0344	0.0606	0.0341	0.0606	0.0395	0.0620
	Medium	γ_1	0.005	0.3643	0.0048	0.3652	0.0046	0.3648
		γ_2	-0.0153	0.1388	-0.0159	0.1397	-0.0102	0.1454
		γ_3	-0.0198	0.1289	-0.0161	0.1286	-0.0071	0.1343
	Large	γ_1	0.0153	0.1704	0.0148	0.1708	0.0155	0.2311
		γ_2	0.1532	0.1229	0.1518	0.1234	0.1540	0.1406
		γ_3	-0.1518	0.1074	-0.1435	0.1065	-0.1477	0.1251
30	Small	γ_1	0.0002	0.0474	0.0003	0.0471	0.0003	0.0479
		γ_2	-0.012	0.0417	-0.0119	0.042	-0.0147	0.0426
		γ_3	-0.0166	0.0318	-0.0168	0.0318	-0.0112	0.0327
	Medium	γ_1	0.0010	0.1994	0.0009	0.1992	0.0009	0.0411
		γ_2	-0.0116	0.0727	-0.0142	0.0730	-0.0119	0.0060
		γ_3	-0.0235	0.0725	-0.0182	0.0726	-0.0346	0.0058
	Large	γ_1	-0.0122	0.1014	-0.0122	0.1016	-0.0068	0.0904
		γ_2	0.1027	0.0722	0.1024	0.0719	0.1113	0.0791
		γ_3	-0.1318	0.0665	-0.1295	0.0662	-0.1440	0.0685
75	Small	γ_1	-0.0007	0.0381	-0.0007	0.0382	-0.0008	0.0378
		γ_2	-0.0027	0.0249	-0.0026	0.0250	-0.0024	0.0253
		γ_3	0.0013	0.0251	0.0009	0.0252	-0.0018	0.0248
	Medium	γ_1	0.0047	0.0109	0.0046	0.1038	0.0070	0.0128
		γ_2	-0.0152	0.0027	-0.0160	0.0519	-0.0141	0.0030
		γ_3	-0.0303	0.0021	-0.0287	0.0455	-0.0333	0.0022
	Large	γ_1	-0.0002	0.0516	-0.0002	0.0515	-0.0054	0.0688
		γ_2	-0.0473	0.0440	-0.0469	0.0440	-0.0416	0.0442
		γ_3	-0.0153	0.0399	-0.0159	0.0400	-0.0145	0.0413
	Average	**Γ**	0.0308	0.0816	0.0300	0.0885	0.0312	0.0758

of SN are slightly smaller than of S-*t*. Note that for small-**D** and T=75 the average *RelBias* over all the three parameters of the normal prior is smaller than that of the Skew normal and Skew-*t* prior. For the average *RMSEs* of the three priors the opposite holds, however.

Table 6 and Figures 2 show that in the case of normal data for the linear Quadratic model the average *RelBias* (in absolute value) (over all T, **D**, and the three parameters) of the normal prior (N) is larger than that of the Skew normal
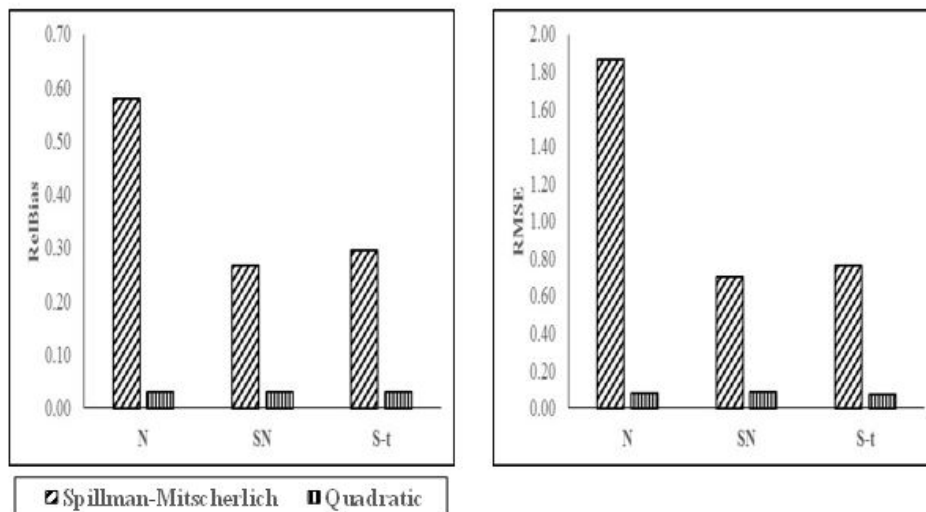
FIGURE 2. Average $RelBias$ and $RMSE$ of the Normal (N), Skew Normal (SN), and Skew-t (S-t) prior for the Spillman-Mitscherlich and linear Quadratic model for normal data

prior (SN), but slightly smaller than that of the Skew-t prior. However, the $RMSE$ of the Skew-t prior is smaller than that of the normal and Skew normal prior. Moreover, the $RMSE$ of the normal prior is smaller than that of the Skew normal prior.

Tables 7 and 8 show the overall fit statistics for the Spillman-Mitscherlich and linear Quadratic model for normal data. For T=10 and all **D** the DIC, EAIC, and EBIC favor the normal prior for both models (with some minor exceptions). For T=30 and 75 all measures favor the Skew-t prior, except for the Spillman-Mitscherlich model, T=30 and small-**D** where the normal prior is selected by all criteria.

6. CONCLUDING REMARKS

This paper analyzed by way of Monte Carlo simulation Bayesian inference of random parameter dose - response models with (i) normal and (ii) skewed, heavy-tailed (Skew-t) distributions of the random response parameter component and independently normally distributed errors. The data generating models were the Skew-t distribution and the normal distribution. The commonly applied linear Quadratic and the nonlinear Spillman-Mitscherlich dose - response model were estimated by means of Bayesian methods. Three priors were considered: a normal, symmetric prior, a Skew normal prior and, finally, a Skew-t prior. The first set of

TABLE 7. *DIC*, *EAIC* and *EBIC* for the Normal (N), Skew-Normal (SN), and Skew-*t* (S-*t*) prior for the Spillman-Mitscherlich model for normal data

(T)	Variance (**D**)	Parameter	Prior		
			N	SN	S-*t*
10	Small	DIC	210.32 (66%)	212.96 (8%)	212.58 (26%)
		EAIC	217.32 (87%)	222.96 (4%)	223.58 (29%)
		EBIC	215.76 (83%)	220.74 (5%)	221.14 (12%)
	Medium	DIC	90.46 (60%)	92.12 (5%)	90.42 (35%)
		EAIC	97.46 (82%)	102.12 (1%)	101.42 (17%)
		EBIC	95.90 (79%)	99.90 (1%)	98.98 (20%)
	Large	DIC	218.21 (81%)	219.85 (3%)	220.27 (16%)
		EAIC	225.21 (99%)	229.85 (0%)	231.27 (1%)
		EBIC	223.66 (99%)	227.63 (0%)	228.83 (1%)
30	Small	DIC	625.59 (65%)	631.07 (0%)	626.16 (35%)
		EAIC	632.59 (82%)	641.07 (0%)	637.16 (18%)
		EBIC	634.38 (85%)	643.63 (0%)	639.97 (15%)
	Medium	DIC	259.90 (3%)	256.85 (8%)	250.02 (89%)
		EAIC	266.90 (15%)	266.85 (5%)	261.02 (80%)
		EBIC	268.69 (22%)	269.40 (5%)	263.83 (73%)
	Large	DIC	687.84 (15%)	691.34 (0%)	683.70 (85%)
		EAIC	694.84 (54%)	701.34 (0%)	694.70 (46%)
		EBIC	696.62 (68%)	703.89 (0%)	697.51 (32%)
75	Small	DIC	1575.76 (0%)	1573.29 (0%)	1560.76 (100%)
		EAIC	1582.76 (4%)	1583.29 (0%)	1571.76 (96%)
		EBIC	1587.33 (8%)	1589.83 (0%)	1578.95 (92%)
	Medium	DIC	661.23 (0%)	647.74 (0%)	632.66 (100%)
		EAIC	668.23 (0%)	657.74 (0%)	643.66 (100%)
		EBIC	672.81 (0%)	664.27 (0%)	650.85 (100%)
	Large	DIC	1764.82 (9%)	1764.62 (9%)	1760.00 (82%)
		EAIC	1771.82 (46%)	1774.62 (1%)	1771.00 (53%)
		EBIC	1776.39 (67%)	1781.15 (0%)	1778.18 (33%)

simulations examined the performance of the three priors in the case of the Skew-*t* data; the second in the case of normal data.

The simulation results showed that the Skew-*t* prior is more accurate and efficient than the normal and Skew-normal in the case of skewed heavy-tailed data. For random components that follow a normal distribution, the Skewed-*t* prior obtain comparable result as the Skew normal prior and more accurate and efficient than the normal in the case of the nonlinear Spillman-Mitscherlich model. For the linear Quadratic model the Skew normal prior is more accurate than the normal and slightly more accurate than the Skew-*t* prior. Furthermore, the Skew-*t* prior

TABLE 8. *DIC*, *EAIC* and *EBIC* for the Normal (N), Skew-Normal (SN), and Skew-*t* (S-*t*) prior for the linear Quadratic model for normal data

			Prior		
(T)	Variance (**D**)	Parameter	N	SN	S-*t*
10	Small	DIC	173.02 (65%)	178.57 (0%)	176.28 (35%)
		EAIC	180.02 (89%)	188.57 (0%)	187.28 (11%)
		EBIC	178.46 (85%)	186.36 (0%)	184.84 (35%)
	Medium	DIC	88.19 (38%)	89.70 (1%)	87.04 (61%)
		EAIC	95.19 (81%)	99.70 (0%)	98.04 (19%)
		EBIC	93.64 (72%)	97.48 (0%)	95.60 (28%)
	Large	DIC	140.08 (41%)	141.82 (4%)	139.02 (55%)
		EAIC	147.08 (80%)	151.82 (0%)	150.02 (20%)
		EBIC	145.52 (75%)	149.60 (0%)	147.58 (25%)
30	Small	DIC	509.04 (19%)	510.21 (0%)	503.82 (81%)
		EAIC	516.04 (48%)	520.21 (2%)	514.82 (50%)
		EBIC	517.82 (53%)	522.76 (0%)	517.63 (47%)
	Medium	DIC	257.53 (0%)	253.68 (3%)	245.17 (97%)
		EAIC	264.53 (9%)	263.68 (5%)	256.17 (86%)
		EBIC	266.32 (12%)	266.23 (5%)	258.98 (83%)
	Large	DIC	411.27 (4%)	409.42 (3%)	401.77 (92%)
		EAIC	418.27 (29%)	419.42 (1%)	412.77 (70%)
		EBIC	420.06 (38%)	421.97 (0%)	415.57 (62%)
75	Small	DIC	1266.50 (3%)	1261.21 (0%)	1245.83 (97%)
		EAIC	1273.50 (3%)	1271.21 (0%)	1256.83 (97%)
		EBIC	1278.07 (5%)	1277.75 (0%)	1264.01 (95%)
	Medium	DIC	647.19 (0%)	633.57 (0%)	615.52 (100%)
		EAIC	654.19 (0%)	643.57 (0%)	626.52 (100%)
		EBIC	658.76 (0%)	650.11 (0%)	633.71 (100%)
	Large	DIC	1026.05 (0%)	1021.44 (2%)	1004.22 (98%)
		EAIC	1033.05 (1%)	1031.44 (1%)	1015.22 (98%)
		EBIC	1037.62 (3%)	1037.97 (1%)	1022.41 (96%)

is more efficient than the normal and Skew normal prior. Overall, the Skew-*t* prior seems to be preferable to the normal and Skew-normal alternatives for dose response modeling, especially because skewed response data is more common than normal response data and the linear Quadratic model is preferable to the nonlinear Spillman-Mitscherlich model in many cases.

REFERENCES

[1] Arellano-Valle, R. B., Bolfarine, H., and Lachos, V. H., Skew-normal linear mixed models, *J. Data Sci.*, **3**(2005), 415-438.

[2] Arellano-Valle, R. B., Bolfarine, H., and Lachos, V. H., Bayesian inference for Skew-normal linear mixed models, *J. Appl. Stat.*, **3**(2007), 663-682.

[3] Bandyopadhyay, D., Lachos, V. H., Castro, L. M., and Dey, D., "Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads", *Biom. J.* **54**(2012), 405-425.

[4] Bandyopadhyay, D., Castro, L. M., Lachos, V. H., and Pinheiro, H. P., "Robust joint nonlinear mixed-effects models and diagnostics for censored HIV viral loads with CD4 measurement error", *J. Agric. Biol. Environ. Stat.* **20**(2015), 121-139.

[5] Boyer, C.N., Larson, J.A., Roberts, R.K., McClure, A.T., Tyler, D.D., and Zhou, V., Stochastic corn yield response functions to nitrogen for corn after corn, corn after cotton, and corn after soybeans, *J. Agric. Appl. Econ.*, **45**(2013), 669-681.

[6] Brorsen, B.W., Using Bayesian estimation and decision theory to determine the optimal level of nitrogen in cotton, Selected Paper, Southern Agricultural Economics Association, Orlando, Florida, (2013).

[7] de Souza, F. A., Malheiros, E. B., and Carneiro, P. R. O., Positioning and number of nutritional levels in dose-response trials to estimate the optimal-level and the adjustment of the models, *Cincia Rural, Santa Maria*, **44**(2014), 1204-1209.

[8] Gelman, A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A, and Rubin, D.B., *Bayesian Data Analysis*, Chapman Hall/CRC, New York, 2014.

[9] Hagenbuch, N., *A comparison of four methods to analyse a non-linear mixed-effects model using simulated pharmacokinetic data*, Master Thesis, Department of Mathematics, Swiss Federal Institute of Technology Zurich, 2011.

[10] Harring, J. R. and Liu J., "A comparison of estimation methods for nonlinear mixed effects models under model misspecification and data sparseness: a simulation study", *Journal of Modern Applied Statistical Methods*, **15**(2016), 539-569.

[11] Jara, A., Quintana, F., San Martin, E., Linear mixed models with skew-elliptical distributions: a Bayesian approach, *Comput. Statist. Data Anal.*, **52**(2008), 5033-5045.

[12] Kery, M., *Introduction to WinBUGS for Ecologists: A Bayesian approach to regression, ANOVA, mixed models and related analyses*, Academic Press, Amsterdam, The Netherlands, 2010.

[13] Lachos, V. H., Dey, D. K., Cancho, V. G., Robust linear mixed models with skew-normal independent distributions from a Bayesian perspective, *J. Statist. Plann. Inference*, **139**(2009), 4098-4110.

[14] Lachos, V. H., Ghosh P., and Arellano-Valle, R. B., Likelihood based inference for skew-normal independent linear mixed models, *Statist. Sinica.*, **20**(2010), 303-322.

[15] Lachos, V. H., Castro L. M., and Dey D. K., Bayesian inference in nonlinear mixed-effects models using normal independent distributions, *Comput. Statist. Data Anal.*, **64**(2013), 237-252.

[16] Lange, K. and Sinsheimer, J., Normal/independent distributions and their applications in robust regression, *J. Comput. Graph. Stat.*, **2**(1993), 175-198.

[17] Lopez-Bellido, R.J., Castillo, J. E., and Lopez-Bellido, L., Comparative response of bread and durum wheat cultivars to nitrogen fertilizer in a rainfed Mediterranean environment: soil nitrate and N uptake and efficiency, *Nutr. Cycl. Agroecosyst.*, **80**(2008), 121-130.

[18] Makowski, D., Wallach, D., and Meynard, J.M., Statistical methods for predicting the responses to applied N and for calculating optimal N rates, *Agron. J.*, **93**(2001), 531-539.

[19] Makowski, D. and Wallach, D., It pays to base parameter estimation on a realistic description of model errors, *Agronomie*, **22**(2002), 179-189.

[20] Makowski, D. and Lavielle, M., Using SAEM to estimate parameters of response to applied fertilizer, *J. Agric. Biol. Environ. Stat.*, **11**(2006), 45-60.

[21] Ouedraogo, F. B. and Brorsen, B. W., *Bayesian estimation of optimal nitrogen rates with a non-normally distributed stochastic plateau function*, The Southern Agricultural Economics Association (SAEA) Annual Meeting, Dallas, Texas, 2014.

[22] Park, S.C., Brorsen, B.W., Stoecker, A.L. and Hattey, J.A., Forage response to swine effluent: a Cox nonnested test of alternative functional forms using a fast double bootstrap, *J. Agr. Appl. Econ.*, **44**(2012), 593-606.

[23] Pinheiro, J., Bornkamp, B., Glimm, E., and Bretz, F., Model-based dose finding under model uncertainty using general parametric models, *Stat. Med.*, **33**(2014), 1646-1661.

[24] Plan, E. L., Maloney, A., Mentr, F., Karlsson, M. O., and Bertrand J., Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for doseresponse models, *The AAPS J.*, **14**(2012), 420-432.

[25] Plummer, M., *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*, Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria, 2003.

[26] Rizzo, M. L., Statistical Computing with R, Chapman Hall/CRC, New York, 2008.

[27] Rosa, G. J. M., Padovani, C. R., and Gianola, D., Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation, *Biom. J.*, **45**(2003), 573-590.

[28] Sain, G. E., and Jauregui, M.A., Deriving fertilizer recommendations with a flexible functional form, *Agron. J.*, **85** (1993), 934-937.

[29] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **64**(2002), 583-639.

[30] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A., The deviance information criterion: 12 years on, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **76**(2014), 485-493.

[31] Su, Y.S., and Yajima, M., *R2jags: A package for running jags from R*, R package version 0.5-7, 2015.

[32] Tanner, M. A., *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag New York, 1993.

[33] Tembo, G., Brorsen, B. W., Epplin, F. M., and Tostao, E., Crop input response functions with stochastic plateaus, *Amer. J. Agr. Econ.*, **90**(2008), 424-434.

[34] Tumusiime, E., Brorsen, B.W., Mosali, J., Johnson, J., Locke, J. and Biermacher, J.T., Determining optimal levels of nitrogen fertilizer using random parameter models, *J. Agr. Appl. Econ.*, **43**(2011), 541-552.

[35] Wallach, D., Regional optimization of fertilization using a hierarchical linear model, *Biometrics*, **51**(1995), 338-346.

[36] Ward, E. J., A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools, *Ecological Modelling*, **211**(2008), 110.

[37] WHO, *Principles for Modelling Dose-Response For The Risk Assessment of Chemicals*, WHO Press, World Health Organization, Geneva, Switzerland, 2009.

Appendix A. R program of Gibbs sampler for Bivariate distribution.

Let (y_1, y_2) be a single observation from a bivariate normally distributed population with unknown mean $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and covariance matrix $\begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_1^2 \end{pmatrix}$. With a uniform prior distribution on $\boldsymbol{\theta}$, the posterior distribution is

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} | y \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_1^2 \end{pmatrix} \right).$$

Although it is simple to draw directly from the joint posterior distribution of (θ_1, θ_2), we consider the Gibbs sampler for the purpose of illustration. To apply the Gibbs sampler to (θ_1, θ_2), we need the conditional posterior distributions which are

$$\theta_1|\theta_2, y \sim N(y_1 + \rho\tfrac{\sigma_1}{\sigma_2}(\theta_2 - y_2), (1 - \rho^2)\sigma_1^2)$$
$$\theta_2|\theta_1, y \sim N(y_2 + \rho\tfrac{\sigma_2}{\sigma_1}(\theta_1 - y_1), (1 - \rho^2)\sigma_2^2)$$

The Gibbs sampler proceeds by alternately sampling from these two normal distribution. For a bivariate distribution (θ_1, θ_2), the Gibbs sampler algorithm is as follows

1. Initialize θ_0 at time $t = 0$.
2. For each iteration, indexed $t = 1, 2, ...$ do:
    a. Set $\theta_{t-1} = (\theta_1, \theta_2)$
    b. Generate $\theta_{1(1)}$ from $f(\theta_1|\theta_2)$
    c. Update $\theta_1 = \theta_{1(1)}$
    d. Generate $\theta_{2(1)}$ from $f(\theta_2|\theta_1)$
    e. Set $\theta_t = (\theta_{1(t)}, \theta_{2(t)})$

```
#Let X=θ, x1=θ1, x2=θ2
#initialize constants and parameters
N < − 5000 #length of chain
burn < − 1000 #burn-in length
X < − matrix(0, N, 2) #the chain, a bivariate sample
rho < − -.75 #correlation
mu1 < − 0
mu2 < − 2
sigma1 < − 1
sigma2 < − .5
s1 < − sqrt(1-rho^2)*sigma1
s2 < − sqrt(1-rho^2)*sigma2
# generate the chain
X [1, ] < − c (mu1, mu2) #initialize
for (i in 2:N) {
x2 < − X [i-1, 2]
m1 < − mu1 + rho * (x2 - mu2) * sigma1/sigma2
X [i, 1] < − rnorm (1, m1, s1)
x1 < − X [i, 1]
m2 < − mu2 + rho * (x1 - mu1) * sigma2/sigma1
X [i, 2] < − rnorm (1, m2, s2) }
b < − burn + 1
x < − X [b:N, ]
#compare sample statistics to parameters
colMeans(x)
cov(x)
cor(x)
plot(x, main=, cex=.5, xlab=bquote(X[1]), ylab= bquote(X[2]), ylim=range(x[,2]))
```