Journal of Multivariate Analysis 135 (2015) 163–174

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models

Haydar Demirhan [a], Zeynep Kalaylioglu [b,*]

[a] Department of Statistics, Hacettepe University, Beytepe, 06800 Ankara, Turkey
[b] Department of Statistics, Middle East Technical University, 06800 Ankara, Turkey

ARTICLE INFO

Article history:
Received 12 August 2013
Available online 12 January 2015

AMS subject classification:
62F15

Keywords:
Hierarchical models
Multi-level models
Multivariate log gamma
Random coefficient
Random effect
Variance components
Hyperprior
Hyperparameter
Directional derivative
Sensitivity analysis

ABSTRACT

In random effect models, error variance (stage 1 variance) and scalar random effect variance components (stage 2 variances) are a priori modeled independently. Considering the intrinsic link between the stages 1 and 2 variance components and their interactive effect on the parameter draws in Gibbs sampling, we propose modeling the variances of the two stages a priori jointly in a multivariate fashion. We use random effects linear growth model for illustration and consider multivariate distributions to model the variance components jointly including the recently developed generalized multivariate log gamma (G-MVLG) distribution. We discuss these variance priors as well as the independent variance priors exercised in the literature in different aspects including noninformativeness and propriety of the associated posterior density. We show through an extensive simulation experiment that modeling the variance components of different stages multivariately results in better estimation properties for the response and random effect model parameters compared to independent modeling. We scrutinize the sensitivity of response model coefficient estimates to the parameters of considered noninformative variance priors and find that their full conditional expectations are insensitive to noninformative G-MVLG prior parameters. We apply independent and joint models for analysis of a real dataset and find that multivariate priors for variance components lead to better fitted hierarchical model than the univariate variance priors.

1. Introduction

Hierarchical models are extensively used to model response data obtained from repeated measures designs, longitudinal studies, and multi-level randomized experiments designed in latin square, split plot, balanced/imbalanced block with random effects. Random effects models are currently very popular in a wide variety of fields such as medicine, pharmacology, psychology, regional sciences, agriculture, sports, modeling of traffic accidents, and energy economy [23,14,16,24,17,5,1,8,18,9].

In a hierarchical model, regression coefficients or treatment effects are viewed as random variables. The top stage (stage 1) of a hierarchical model consists of the response model whereas the next stage (stage 2) consists of models for the random coefficients (random effects). For responses obtained from a repeated measures design or a longitudinal study, the random coefficients of a linear hierarchical model account for the heterogeneity among the subjects as well as the correlation among the observations collected from the same subject at different time points. For data obtained from a randomized experiment in which the groups are viewed as a random selection from a population of groups, random effects encapture group specific effects as well as between group variation. For Bayesian analysis of hierarchical models, the hierarchical

* Corresponding author.
E-mail addresses: haydarde@hacettepe.edu.tr (H. Demirhan), kzeynep@metu.edu.tr (Z. Kalaylioglu).

http://dx.doi.org/10.1016/j.jmva.2014.12.013

structure is enlarged to include yet another stage at which the variances of the random coefficients (random effects) are given prior distributions. This stage is the focus of the current article.

As there is usually and unsurprisingly no sufficient prior knowledge regarding what could be the variance of the random coefficient, the user prefers noninformative hyperpriors and let the likelihood dominate the inference on the stage 2 variances. Therefore of interest are the diffuse priors and researchers in the area have been in quest for what could be regarded as default hyperprior for the stage 2 variance parameters or a one to one transformation of them. Of all the diffuse priors considered in the literature, gamma distribution with small shape and scale parameters (denoted thereof by $Ga(\epsilon, \epsilon)$) has been the most commonly used default prior for the inverse of stage 2 variance parameter (an equivalent representation being $Inv - Ga(\epsilon, \epsilon)$ for stage 2 variance) owing its common use to its conjugacy to Normality and resulting computational benefits in softwares such as BUGS that perform Gibbs sampling for posterior inference. A dangerous but often overlooked characterization of diffuse hyperprior distributions based on gamma distribution is that it may result in near or complete improper posteriors. For instance, Natarajan and McCulloch [21] discuss diffuse inverted gamma priors in probit-Normal hierarchical models resulting in improper posterior distributions and inaccurate posterior estimates. Motivated for developing proper hyperprior, Natarajan and Kass [20] proposed for generalized linear mixed models an approximate uniform shrinkage and Jeffreys priors for the unstructured second stage variance matrix and showed that their priors lead to proper posteriors and have better frequentist properties relative to inverse-gamma and Wishart hyperpriors.

More recently Lambert et al. [15] compare effects of 13 different prior settings induced on stage 2 scale parameters of a random effects hierarchical model via a simulation study using WinBUGS. They consider various gamma, Pareto and logistic distributions as prior for stage 2 precision, various uniform distributions as prior for stage 2 variance, its square root and natural logarithm, and various half-normal distributions as prior for square root of stage 2 variance. Not a particular prior setting is identified as best in all scenarios and they note that uniform prior is not a good alternative if a vague prior is intended for stage 2 variance.

Browne and Draper [4] for Bayesian analysis of mixed linear and random effects logistic regression models consider $Inv - Ga(\epsilon, \epsilon)$ and uniform prior on $(0, 1/\epsilon)$ for the stage 2 variance. Their simulation study demonstrates that Bayesian interval inference with these priors face undercoverage problems in mixed linear models when the number of level 2 units of the experimental design is small. Gelman [11] considers traditional $Inv - Ga(\epsilon, \epsilon)$ and Uniform(0,A) hyperpriors and constructs a folded-noncentral-t family of priors as hyperpriors for variance parameters in hierarchical models. Unlike Lambert et al. [15], Gelman [11] suggests the use of uniform prior for a noninformative prior setting. He recommends half-Cauchy (denoted thereof by $HC(0, 1)$) distribution, which is included in folded-noncentral-t family, as weakly informative prior for stage 2 standard deviation and advises not to use the inverse-gamma setting. Of these prior distributions, as indicated in the article $\lim_{A\to\infty}$ Uniform(0,A) yields proper posterior whereas $\lim_{\epsilon\to 0} Ga(\epsilon, \epsilon)$ does not and the posterior inference is sensitive to the choice of ϵ.

Polson and Scott [22] propose to induce half-Cauchy distribution on stage 2 standard deviation and obtained inverted-beta priors for stage 2 variance which ultimately led to the class of hypergeometric inverted-beta distributions resulting in a generalization of the half-Cauchy prior. They qualify the half-Cauchy prior as a sensible default prior for scale parameters in hierarchical models.

One should note, however, that there are two main aspects with these priors that need attention. First, with these priors, variance components of different stages are a priori modeled independently although they are linked as they are the components of the total variation in a response. Second, as presented in Section 3.2, the drawback of these prior structures is that the posterior inference on the response model coefficients in a hierarchical model is highly sensitive to the choice of the parameters of these prior distributions. In this article we a priori model the variance components of different stages jointly by specifying a multivariate prior distribution. Desirable properties of such a joint variance prior density are 1. noninformative, 2. leads to proper posteriors, and 3. change in the parameters of the variance priors do not effect the posterior inference on response model coefficients.

For joint prior modeling, we stack stage 1 and stage 2 variances and induce a multivariate hyperprior distribution. We consider multivariate normal, multivariate skew normal, and generalized multivariate log-gamma distribution as the multivariate hyperprior distribution on natural logarithms of the variance components and investigate their properties based on our prototype hierarchical model.

The rest of the article is organized as follows. In Section 2, we discuss certain modeling aspects concerning the variance components including the informativeness issue, present the proposed joint variance prior setting, and discuss its propriety. Section 3 presents an extensive simulation study in which we investigate and compare sensitivity of the posterior estimators of the proposed joint prior to those in the literature where variances of different stages are a priori modeled independently. In this section, the notion of noninformativeness for a multivariate prior density is furnished and subspace of variance hyperparameters to which the posterior inference is rather insensitive is sought through the directional derivative concept. A data application is presented in Section 4. Finally a discussion on the evaluation of the results and generalization of the proposed approach for further modeling extensions is given in Section 5.

2. Modeling the variance components

We will consider the basic random coefficient model given in (1). Such basic models are also considered in Bayesian literature to study variance components in normal hierarchical models [11,22]. The model is

$$
\begin{aligned}
&Y_{ij} = \alpha_i + \beta_i(x_j - \bar{x}) + e_{ij},\\
&\alpha_i \sim N(\alpha_c, \tau_\alpha),\\
&\beta_i \sim N(\beta_c, \tau_\beta),
\end{aligned} \tag{1}
$$

and e_{ij}'s are iid $N(0, \tau_c)$ for $i = 1, \ldots, n, j = 1, \ldots, J$. Additional set of assumptions is $Cov(e_{ij}, \alpha_k) = 0$, $Cov(e_{ij}, \beta_k) = 0$, and $Cov(\alpha_i, \beta_i) = 0$ for all i, j, and k. This model is basically a random effects linear growth model, an important class of hierarchical models. Here, the variance components τ_c and $(\tau_\alpha, \tau_\beta)^T$ constitute stage 1 and stage 2 variances, respectively. Although this is a simple response model, it is not a problem as our focus is on the variance priors. This response model is only used to illustrate our approach for variance prior modeling. The proposed approach is independent of how the response is modeled and therefore can easily be adopted for more complex response model extensions.

In this section, we present the traditional univariate and proposed multivariate priors for the variance components using the prototype model described in Section 1. The Bayesian hierarchical configuration of model (1) is as follows:

$$
\begin{aligned}
&Y_{ij}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{\tau} \overset{ind}{\sim} N(\alpha_i + \beta_i(x_j - \bar{x}), \tau_c),\\
&\alpha_i \sim N(\alpha_c, \tau_\alpha),\ \beta_i \sim N(\beta_c, \tau_\beta),\\
&\alpha_c \sim N(\mu_\alpha, \varpi_\alpha^2),\ \beta_c \sim N(\mu_\beta, \varpi_\beta^2),\\
&\tau_c \sim F_c(\cdot),\\
&\tau_\alpha \sim F_\alpha(\cdot),\ \tau_\beta \sim F_\beta(\cdot),
\end{aligned} \tag{2}
$$

where $\boldsymbol{\tau} = (\tau_\alpha, \tau_\beta, \tau_c)^T$, $F_c(\cdot)$, $F_\alpha(\cdot)$ and $F_\beta(\cdot)$ are used to denote the prior distributions of variance components. The last two lines of this model are the focus of this article. Below Θ symbolizes the vector of all the parameters in the resulting Bayesian hierarchical models. Also if e.g. θ is a symbol for a parameter in Θ, then $\Theta_{-\theta}$ refers to all the parameters in Θ excluding the θ itself.

2.1. Traditional variance priors

Inverse-gamma model: The most practical approach for modeling the variance parameters has been to induce inverse-gamma (IG) priors on τ_α, τ_β and τ_c as in the following:

$$
\begin{aligned}
&\tau_c \sim IG(\xi_c, \eta_c),\\
&\tau_\alpha \sim IG(\xi_\alpha, \eta_\alpha),\ \tau_\beta \sim IG(\xi_\beta, \eta_\beta).
\end{aligned} \tag{3}
$$

We call this prior setting IG model throughout the article. Under the IG model, the full conditional posterior distributions of the parameters are as follows:

$$
\begin{aligned}
&\tau_c|\Theta_{-\tau_c}, \boldsymbol{y} \sim IG(\xi_c + 0.5nJ, \eta_c + 0.5S_4),\\
&\tau_\alpha|\Theta_{-\tau_\alpha}, \boldsymbol{y} \sim IG(\xi_\alpha + 0.5n, \eta_\alpha + 0.5S_\alpha),\\
&\tau_\beta|\Theta_{-\tau_\beta}, \boldsymbol{y} \sim IG(\xi_\beta + 0.5n, \eta_\beta + 0.5S_\beta),\\
&\alpha_c|\Theta_{-\alpha_c}, \boldsymbol{y} \sim N\left[\left(\frac{\mu_\alpha}{\sigma_\alpha^2} + \frac{\sum \alpha_i}{\tau_\alpha}\right)\left(\frac{1}{\sigma_\alpha^2} + \frac{n}{\tau_\alpha}\right)^{-1}, \left(\frac{1}{\sigma_\alpha^2} + \frac{n}{\tau_\alpha}\right)^{-1}\right],\\
&\beta_c|\Theta_{-\beta_c}, \boldsymbol{y} \sim N\left[\left(\frac{\mu_\beta}{\sigma_\beta^2} + \frac{\sum \beta_i}{\tau_\beta}\right)\left(\frac{1}{\sigma_\beta^2} + \frac{n}{\tau_\beta}\right)^{-1}, \left(\frac{1}{\sigma_\beta^2} + \frac{n}{\tau_\beta}\right)^{-1}\right],\\
&\alpha_i|\Theta_{-\alpha_i}, \boldsymbol{y} \sim N\left[\left(\frac{\alpha_c}{\tau_\alpha} + \frac{S_1 - S_2}{\tau_c}\right)\left(\frac{1}{\tau_\alpha} + \frac{J}{\tau_c}\right)^{-1}, \left(\frac{1}{\tau_\alpha} + \frac{J}{\tau_c}\right)^{-1}\right],\\
&\beta_i|\Theta_{-\beta_i}, \boldsymbol{y} \sim N\left[\left(\frac{\beta_c}{\tau_\beta} + \frac{S_1^* - S_2^*}{\tau_c}\right)\left(\frac{1}{\tau_\beta} + \frac{S_3^*}{\tau_c}\right)^{-1}, \left(\frac{1}{\tau_\beta} + \frac{S_3^*}{\tau_c}\right)^{-1}\right],
\end{aligned} \tag{4}
$$

where $S_4 = \sum_{i,j}[y_{ij} - (\alpha_i + \beta_i(x_j - \bar{x}))]^2$, $S_\alpha = \sum_i(\alpha_i - \alpha_c)^2$, $S_\beta = \sum_i(\beta_i - \beta_c)^2$, $S_1 = \sum_{i,j} y_{ij}$, $S_2 = \beta_i \sum_j(x_j - \bar{x})$, $S_1^* = \sum_{i,j} y_{ij}(x_j - \bar{x})$, $S_2^* = \alpha_i \sum_j(x_j - \bar{x})$ and $S_3^* = \sum_j(x_j - \bar{x})^2$. Noting that variance of a random variable distributed as IG is $\frac{\eta^2}{(\xi-1)^2(\xi-2)}$, shape (ξ) and scale (η) parameters should be set close to 2 and at a large value respectively for a noninformative IG prior with positive variance. The application of diffuse IG model in the literature has been different: IG model is applied such that $Ga(\epsilon, \epsilon)$, where ϵ is a small value e.g. 0.001, is assumed as prior density for $1/\tau_.$. For instance, $1/\tau_. \sim Ga(0.001, 0.001)$ which in turn is equivalent to $\tau_. \sim IG(0.001, 1000)$. This use has two problems worthy of consideration: 1. limiting $Ga(\epsilon, \epsilon)$ is improper as $\epsilon \to 0$ and improper prior in this case leads to improper posterior, 2. variance of the resulting IG parameter is negative! The former one is about the prior density becoming improper in the limit of noninformativeness. The later situation seems to escape from computational problems while Gibbs sampling, however it is contradicting with the basic property of variance. Direct use of noninformative IG on the variance component itself as described above removes these problems.

Uniform model: The second prior of concern is inducing uniform priors on τ_α and τ_β, and an inverse-gamma prior on τ_c as follows

$$\tau_c \sim IG(\xi_c, \eta_c),$$
$$\tau_\alpha \sim U(0, a),\ \tau_\beta \sim U(0, b),$$

where $a > 0$ and $b > 0$. The model with these variance priors is referred as UNF model throughout the article. For UNF model, full conditional posterior distributions of $\tau_c, \alpha_c, \beta_c, \alpha_i$ and β_i are the same as in (4) along with the full conditional posterior distributions of τ_α and τ_β given as

$$\tau_\alpha|\Theta_{-\tau_\alpha}, \boldsymbol{y} \sim IG(0.5n - 1, 0.5S_\alpha)I(\tau_\alpha < a),$$
$$\tau_\beta|\Theta_{-\tau_\beta}, \boldsymbol{y} \sim IG(0.5n - 1, 0.5S_\beta)I(\tau_\beta < b).$$

Noninformative priors for τ_α and τ_β are obtained by setting a and b at large values.

Half-Cauchy model: Third we consider half-Cauchy (HC) priors on both τ_α and τ_β, and an inverse-gamma prior on τ_c as follows

$$\tau_c \sim IG(\xi_c, \eta_c),$$
$$\tau_\alpha \sim HC(0, 1),\ \tau_\beta \sim HC(0, 1).$$

This model is called HC model throughout the article. For the HC model, full conditional posterior distributions of $\tau_c, \alpha_c, \beta_c$, α_i and β_i are the same as in IG model while those for τ_α and τ_β are obtained as the following:

$$p(\tau_\alpha|\Theta_{-\tau_\alpha}, \boldsymbol{y}) \propto \tau_\alpha^{-n/2}(1 + \tau_\alpha^2)^{-1} \exp\{-0.5S_\alpha/\tau_\alpha\},\ \tau_\alpha > 0, \tag{5}$$

and

$$p(\tau_\beta|\Theta_{-\tau_\beta}, \boldsymbol{y}) \propto \tau_\beta^{-n/2}(1 + \tau_\beta^2)^{-1} \exp\{-0.5S_\beta/\tau_\beta\},\ \tau_\beta > 0. \tag{6}$$

2.2. Joint variance priors

Following facts motivate modeling τ_c, τ_α, and τ_β jointly. First, τ_α, τ_β, and τ_c are intrinsically linked as they are the components of the total variation in an observed response, e.g. for model (1) total variation in a response y_{ij} is equal to $\tau_\alpha + x_j^2\tau_\beta + \tau_c$. Second, the conditional probability of τ_β given all the other model parameters involves τ_α and τ_c. Similar situation holds for τ_α. These facts altogether motivate joint modeling of $(\tau_c, \tau_\alpha, \tau_\beta)^T$. Fully noninformative approach in multivariate modeling of variance priors has two dimensions in need of consideration; diffuseness of the prior distribution and the correlation structure among the parameters. Noninformativeness in the case of a multivariate prior distribution is associated with large variances and low correlations. Variances should be set high to reflect low degree of belief in prior information. One should have strong knowledge in order to be able to assign large correlations between the parameters and thus low correlations should be induced for a noninformative setting. We consider the following multivariate hyperprior models for the variance components.

Generalized multivariate log-gamma model: We first consider generalized multivariate log-gamma (G-MVLG) distribution developed by Demirhan and Hamurkaroglu [7]. Accordingly, if $\boldsymbol{W} \sim G-MVLG(\delta, \nu, \boldsymbol{\lambda}^T, \boldsymbol{\eta}^T)$, then the joint probability density function (pdf) of $\boldsymbol{W} = (W_1, \ldots, W_k)$ is as follows

$$p(\boldsymbol{w}|\boldsymbol{g}) \propto \delta^\nu \sum_{m=0}^{\infty} \frac{(1-\delta)^m \prod_{j=1}^{k} \eta_j \lambda_j^{-\nu-m}}{[\Gamma(\nu+m)]^{k-1}\Gamma(\nu)m!} \exp\left\{(\nu+m)\sum_{j=1}^{k}\eta_j w_j - \sum_{j=1}^{k}\frac{1}{\lambda_j}\exp\{\eta_j w_j\}\right\},$$

where $\boldsymbol{w} \in \mathbb{R}^k$, $\nu > 0$, $\lambda_j > 0$, $\eta_j > 0$, $\boldsymbol{\eta} = (\eta_j)$, $\boldsymbol{\lambda} = (\lambda_j)$, for $j = 1, \ldots, k$, $\delta = \det(\boldsymbol{\Omega})^{\frac{1}{k-1}}$,

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \sqrt{\text{abs}(\rho_{12})} & \cdots & \sqrt{\text{abs}(\rho_{1k})} \\ \sqrt{\text{abs}(\rho_{12})} & 1 & \cdots & \sqrt{\text{abs}(\rho_{2k})} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\text{abs}(\rho_{1k})} & \sqrt{\text{abs}(\rho_{2k})} & \cdots & 1 \end{pmatrix}, \tag{7}$$

$\det(\cdot)$ and $\text{abs}(\cdot)$ respectively denote the determinant and absolute value of the inner expression, ρ_{ij} stands for the correlation between ith and jth random components, and $\boldsymbol{g} = (\delta, \nu, \boldsymbol{\lambda}^T, \boldsymbol{\eta}^T)$ is the vector of parameters characterizing the distribution.

Marginal expectations and variances are given as follows

$$E(W_j) = \frac{1}{\eta_j}\left[\ln(\lambda_j/\delta) + \psi(\nu)\right] \text{ and } Var(W_j) = \psi^{[1]}(\nu)/(\eta_j)^2, \tag{8}$$

where $\psi(\cdot)$ and $\psi^{[1]}(\cdot)$ are digamma and trigamma functions, respectively. As seen in Eq. (8), η_j and λ_j are influential on the marginal variances and expected values, respectively. Impact of correlations between variables is reflected by δ. Effects of η_j, λ_j, δ and ν over the marginal expectations and variances are discussed in detail by Demirhan and Hamurkaroglu [7].

Let log transformed variance parameters be $\theta_1 = \log(\tau_\alpha)$, $\theta_2 = \log(\tau_\beta)$ and $\theta_3 = \log(\tau_c)$. Then prior distribution of the variance components in our joint hierarchical model, which is referred as MVLG model hereafter, is as follows:

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim G - MVLG(\delta, \nu, \boldsymbol{\lambda}^T, \boldsymbol{\eta}^T). \tag{9}$$

Full conditional posterior distributions of α_c and β_c under MVLG model are the same as in (4), and those for the rest of the parameters and the hyperparameters are obtained as follows:

$$\begin{aligned} p(\boldsymbol{\theta}|\Theta_{-\boldsymbol{\theta}}, \boldsymbol{y}) \;\propto\; & \exp\left\{\sum_{j=1}^{2}\left[-(0.5n-1)+\nu\eta_j\right]\theta_j + \left[-(0.5nJ-1)+\nu\eta_3\right]\theta_3\right\} \\ & \times \exp\left\{S_5\prod_{j=1}^{3}e^{\eta_j\theta_j} - \sum_{j=1}^{3}\lambda_j^{-1}e^{\eta_j\theta_j}\right\} \times \exp\left\{-0.5S_\alpha e^{-\theta_1} - 0.5S_\beta e^{-\theta_2} - 0.5S_4 e^{-\theta_3}\right\}, \quad \boldsymbol{\theta} \in \mathbb{R}^3 \\ \alpha_i|\Theta_{-\alpha i}, \boldsymbol{y} \sim\; & N\left[\left(\frac{\alpha_c}{e^{\theta_1}} + \frac{S_1 - S_2}{e^{\theta_3}}\right)\left(\frac{1}{e^{\theta_1}} + \frac{J}{e^{\theta_3}}\right)^{-1}, \left(\frac{1}{e^{\theta_1}} + \frac{J}{e^{\theta_3}}\right)^{-1}\right], \quad \alpha_i \in \mathbb{R} \\ \beta_i|\Theta_{-\beta i}, \boldsymbol{y} \sim\; & N\left[\left(\frac{\beta_c}{e^{\theta_2}} + \frac{S_1^* - S_2^*}{e^{\theta_3}}\right)\left(\frac{1}{e^{\theta_2}} + \frac{S_3^*}{e^{\theta_3}}\right)^{-1}, \left(\frac{1}{e^{\theta_2}} + \frac{S_3^*}{e^{\theta_3}}\right)^{-1}\right], \quad \beta_i \in \mathbb{R}, \end{aligned} \tag{10}$$

where $S_5 = (1-\delta)\exp\{-2.207728\}\prod_{j=1}^{3}\lambda_j$. Derivation of $p(\theta_1, \theta_2, \theta_3|\boldsymbol{y})$ is given in Appendix A.

For a noninformative MVLG model, one should assign small values to the correlations between parameters (ρ_{ij}) and near zero values to η_j. It can be seen straightforwardly that $p(\boldsymbol{w}|\boldsymbol{g})$ does not become improper as $\eta_j \to 0$.

Multivariate normal model: Second, we consider multivariate normal (MVN) distribution to jointly model the log transformed variance parameters, namely $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim MVN(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$. Full conditional posterior distributions of α_c and β_c under MVN model are the same as in (4), those of α_i and β_i are the same as given in (10) and full conditional posterior distribution of $\boldsymbol{\theta}$ is straightforwardly obtained as

$$\begin{aligned} p(\boldsymbol{\theta}|\Theta_{-\boldsymbol{\theta}}, \boldsymbol{y}) \propto \exp\Big\{ & -(0.5n-1)(\theta_1+\theta_2) - (0.5nJ-1)\theta_3 - 0.5S_\alpha e^{-\theta_1} - 0.5S_\beta e^{-\theta_2} \\ & - 0.5S_4 e^{-\theta_3} - 0.5(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)\boldsymbol{\Sigma}_\theta^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta)^T\Big\}, \quad \boldsymbol{\theta} \in \mathbb{R}^3. \end{aligned} \tag{11}$$

In order to get a noninformative setting for MVN model, first large values should be chosen for the diagonal elements of $\boldsymbol{\Sigma}_\theta$, then values for off-diagonal elements of $\boldsymbol{\Sigma}_\theta$ (i.e. the covariances) should be selected so that the correlations between the parameters are low while at the same time ensuring the positive definiteness of $\boldsymbol{\Sigma}_\theta$.

Multivariate skew normal model: We also consider multivariate skew normal (MVSN) distribution for the joint prior distribution of the log transformed variance parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim MVSN(\boldsymbol{\lambda}, \boldsymbol{\Psi})$. Reader should refer to Azzalini and Valle [3] and Azzalini [2] for MVSN distribution. $\boldsymbol{\Psi}$ is the dependence matrix and the vector $\boldsymbol{\lambda} = (\xi_i(1-\xi_i^2)^{-0.5})$ where $\xi_i \in (-1, 1)$ and determines skewness. Full conditional posterior distributions of α_c and β_c under MVSN model are the same as in (4), those of α_i and β_i are the same as given in (10) and full conditional posterior distribution of $\boldsymbol{\theta}$ is easily obtained as the following:

$$\begin{aligned} p(\boldsymbol{\theta}|\Theta_{-\boldsymbol{\theta}}, \boldsymbol{y}) \propto \exp\Big\{ & -(0.5n-1)(\theta_1+\theta_2) - (0.5nJ-1)\theta_3 - 0.5S_\alpha e^{-\theta_1} - 0.5S_\beta e^{-\theta_2} \\ & - 0.5S_4 e^{-\theta_3}\Big\}\phi(\boldsymbol{\theta}, \boldsymbol{\Omega})\Phi(\boldsymbol{\gamma}^T\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \mathbb{R}^3, \end{aligned} \tag{12}$$

where $\phi(\boldsymbol{\theta}, \boldsymbol{\Omega})$ is the pdf of $MVN(\mathbf{0}, \boldsymbol{\Omega})$, $\Phi(\boldsymbol{\gamma}^T\boldsymbol{\theta})$ is the cdf of $N(0, 1)$, $\boldsymbol{\Omega} = \boldsymbol{\Delta}(\boldsymbol{\Psi} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T)\boldsymbol{\Delta}$, for $i = 1, 2, 3$, $\boldsymbol{\Delta} = \mathrm{diag}((1-\xi_i^2)^{0.5})$, and $\boldsymbol{\gamma} = (\boldsymbol{\lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Delta}^{-1})(1+\boldsymbol{\lambda}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda})^{-0.5}$.

For a noninformative MVSN model, $\boldsymbol{\Psi}$ and vector $\boldsymbol{\xi}$ are assigned small values in order to make the distribution slightly skewed in a noninformative setting.

3. Influence of variance priors on posterior inference

In this section, we evaluate and compare influence of different modeling strategies for hierarchical variance components on posterior inferences of main model parameters such as the response model coefficients. An extensive simulation study is conducted to assess and compare accuracy of the posterior estimates obtained from the marginal and the joint prior modeling strategies. An approach utilizing directional derivatives is developed for thorough analytic evaluation of sensitivity

Table 1
Simulation scenarios.

Scenario	J	τ_c	τ_α	τ_β	Scenario	J	τ_c	τ_α	τ_β
1	2	0.5	10	10	13	5	0.5	10	10
2	2	0.5	10	100	14	5	0.5	10	100
3	2	0.5	100	10	15	5	0.5	100	10
4	2	0.5	100	100	16	5	0.5	100	100
5	2	10	10	10	17	5	10	10	10
6	2	10	10	100	18	5	10	10	100
7	2	10	100	10	19	5	10	100	10
8	2	10	100	100	20	5	10	100	100
9	2	100	10	10	21	5	100	10	10
10	2	100	10	100	22	5	100	10	100
11	2	100	100	10	23	5	100	100	10
12	2	100	100	100	24	5	100	100	100

of posterior outcome to the variance hyperparameters under marginal and joint modeling strategies. These endeavors together shed light on (i) effect of different prior constructions for hierarchical variance components on the performance of estimators in normal hierarchical models, (ii) whether the sensitivity of posterior outcome to variance hyperparameters vary over the hyperparameter space.

3.1. Simulation study

The aim of this simulation study is to compare posterior performances of noninformative marginal and joint models. We consider the models given in Section 2 for the hierarchical model given in (1). The simulation study consists of 100 replications. At each iteration of the simulation experiment, the data are generated as follows. For $i = 1, \ldots, 30$, α_i and β_i are generated from $N(0, \tau_\alpha)$ and $N(0, \tau_\beta)$ respectively. For $i = 1, \ldots, 30$ and $j = 1, \ldots, J$, y_{ij} are generated from $N(\alpha_i + \beta_i x_j, \tau_c)$. In order to evaluate the accuracy of posterior estimates under different scenarios, we consider values of various different magnitudes for the variance components τ_c, τ_α, and τ_β as well as number of groups J in designing the simulations. These scenarios are given in Table 1.

The vector $\boldsymbol{x}$ in (1) is taken as $(8, 15)^T$ when $J = 2$ and $(8, 15, 22, 29, 36)^T$ when $J = 5$. In all models and simulations, we set $\mu_\alpha = \mu_\beta = 10^{-3}$ and $\varpi_\alpha^2 = \varpi_\beta^2 = 10^6$. We set $\xi_\alpha = \xi_\beta = \xi_c = 2.1$ and $\eta_\alpha = \eta_\beta = \eta_c = 100$ for a noninformative IG model and $a = b = 100$ for a noninformative UNF model. We consider two settings for the MVLG model that are denoted by MVLG1 and MVLG2 in which the parameter ν is reasonably set at 1.42 based on the fact that its value has no effect on the inferences. The vector $\boldsymbol{\eta}$ is taken as $\boldsymbol{\eta} = (0.01, 0.01, 0.01)^T$ for noninformative MVLG models. Since $\boldsymbol{\lambda}$ has no effect on the noninformativeness of MVLG model, any positive value can be chosen. Following correlation matrices for $(\theta_1, \theta_2, \theta_3)^T$ are considered for the MVLG1 and MVLG2 models respectively and the parameter δ is determined accordingly as described in Section 2.2 based on these correlation matrices.

$$\boldsymbol{\Omega}_1 = \begin{pmatrix} 1 & 0 & 0.1 \\ 0 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{13}$$

For the MVN model, we consider two settings denoted by MVN1 and MVN2 and use $\boldsymbol{\mu}_\theta = (0, 0, 0)^T$ in both settings. The following covariance matrices are used for MVN1 and MVN2 models, respectively

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 100 & 0 & 10 \\ 0 & 100 & 10 \\ 10 & 10 & 100 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{pmatrix}, \tag{14}$$

which give the same correlations between θ_1, θ_2 and θ_3 as in the models MVLG1 and MVLG2, respectively. We consider two different settings for the MVSN model where $\boldsymbol{\xi} = (-0.1, -0.1, -0.1)^T$ for the first one (MVSN1) and $\boldsymbol{\xi} = (0.1, 0.1, 0.1)^T$ for the second one (MVSN2). In both settings $\boldsymbol{\Psi}$ is set equal to $\boldsymbol{\Omega}_1$ of Eq. (13).

For posterior calculations of the IG, UNF and HC models, full conditional posterior distributions of parameters and hyperparameters are analytically derived and used within the Gibbs sampler. For the posterior calculations of MVLG, MNV, and MVSN models, a Metropolis–Hastings step is employed for natural logarithms of stage 2 variance parameters under the Gibbs sampling in which proposals for θ_1, θ_2 and θ_3 are generated from a MVN distribution. For normal proposal densities, one expects to get acceptance rates of about 45%–50% for a univariate case and 25%–30% for a multivariate case to conclude appropriateness of a proposal distribution [6]. In our simulations, proposal distributions are tuned up to make overall acceptance rates about 0.30 for MVLG model and 0.5 for MVN and MVSN models both. Potential scale reduction factor ($\hat{R}$) given by Gelman [10] is used to evaluate the convergence of Gibbs sequences. Suitable burn-in period and thinning for each parameter are determined over pilot runs of chains.

Mean square errors (MSEs) of the estimators in scenarios 15 and 23 are respectively given in Tables 2 and 3. Array of tables corresponding to the rest of the scenarios are found in Appendix B. In the tables, MSE values under $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ columns are averages of the MSEs of α_i and β_i, $i = 1, \ldots, 30$, respectively. Accordingly, in terms of estimation of the main model

Table 2
MSE values for the scenario: ($J = 5$; $\tau_c = 0.5$; $\tau_\alpha = 100$; $\tau_\beta = 10$).

Model	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	τ_α	τ_β	τ_c	α_c	β_c
MVLG1	0.000	8.420	4.760	6.188	4.034	0.03	3.331
MVLG2	0.000	8.418	4.286	8.432	4.361	0.03	3.301
MVSN1	0.224	10.100	>1000	>1000	>1000	0.129	13.60
MVSN2	0.201	8.510	>1000	>1000	>1000	0.104	13.50
MVN1	0.111	7.658	>1000	>1000	>1000	0.02	2.210
MVN2	0.15	7.035	>1000	>1000	>1000	0.024	2.261
IG	0.220	8.744	53.900	>1000	>1000	0.003	0.485
UNF	0.223	8.749	54.000	76.90	>1000	0.007	3.629
HC	>1000	>1000	>1000	>1000	610.0	0.059	1.253

Table 3
MSE values for the scenario: ($J = 5$; $\tau_c = 100$; $\tau_\alpha = 100$; $\tau_\beta = 10$).

Model	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	τ_α	τ_β	τ_c	α_c	β_c
MVLG1	1.003	41.563	7.900	6.182	9.608	0.355	0.336
MVLG2	1.261	41.348	6.689	7.293	5.164	0.356	0.336
MVSN1	33.00	>1000	>1000	>1000	>1000	260.0	>1000
MVSN2	100.0	230.0	>1000	>1000	>1000	240.0	>1000
MVN1	150.0	>1000	>1000	>1000	>1000	140.0	>1000
MVN2	5.072	66.733	>1000	>1000	>1000	0.571	2.409
IG	15.80	110.0	>1000	>1000	200.0	0.884	3.630
UNF	35.63	110.0	150.0	>1000	210.0	0.883	4.680
HC	880.0	94.21	>1000	>1000	>1000	0.894	3.329

parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: MVLG1 and MVLG2 models are superior to the other models based on their MSEs; the models MVN1, MVN2, MVSN1, and MVSN2 have relatively small MSEs when stage 1 variance is small but their MSE increase with the stage 1 variance. In terms of estimation of the variance components τ_α, τ_β, and τ_c: MVLG1 and MVLG2 models produce the smallest MSEs among all the models considered; when the true stage 1 variance is small, modeling a priori the variance components univariately result in unsatisfactory estimation properties which is in parallel with the findings in the literature; performances of MVN1, MVN2, MVSN1 and MVSN2 models deteriorate with larger stage 1 variance. In terms of estimation of the population parameters α_c and β_c: IG, UNF, HC, MVLG1, and MVLG2 models yield small MSEs; MSEs of the models MVN1, MVN2, MVSN1, and MVSN2 lower with increasing stage 1 variance. Overall, the multivariate models MVLG1 and MVLG2 perform the best among all the univariate and multivariate models considered. The rationale behind MVLG1 and MVLG2 performing better than the other multivariate models of the concern is explained as follows. In the models MVLG1 and MVLG2, uncorrelation induced for noninformativeness does not in turn lead to independence of the variance components, that is the essential dependence among the variance components is conveyed with the G-MVLG distribution. On the other hand, the models MVN and MVSN forfeit the intrinsic dependence among the variance components when they are made noninformative. In general, in all simulation combinations, magnitude of the true τ_α and τ_β values have no considerable effect on the MSEs. Univariate modeling approach is negatively affected by small group size while this is not the case for joint modeling approach. Joint modeling with MVSN distribution yields larger MSEs when true stage 1 variance is large.

3.2. Evaluation of posterior sensitivity by directional derivatives

A method based on directional derivatives is tailored to analytically elaborate the effect of a unit change in the variance hyperparameters on the posterior means of response model coefficients, namely $\{\alpha_i, i = 1, \ldots, n\}$ and $\{\beta_i, i = 1, \ldots, n\}$, under marginal and joint modeling strategies. Our strategy is described below for β_i and the same for α_i. Let $(\beta_i^{(1)}, \ldots, \beta_i^{(B)})^T$ be the Markov chain of size B for β_i after the burn-in period. Also, let $\frac{1}{B}\sum_{b=1}^{B}\beta_i^{(b)}$, $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$, and $E_{\beta_i|\boldsymbol{Y}}(\beta_i)$ respectively denote the sample average of the $\beta_i^{(b)}$s, expectation of β_i under the full conditional density, and posterior mean of β_i. To study the aforementioned effect, first $E_{\beta_i|\boldsymbol{Y}}(\beta_i)$ is approximated by $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$ based on the facts that (i) $\frac{1}{B}\sum_{b=1}^{B}\beta_i^{(b)}$ converges almost surely to $E_{\beta_i|\boldsymbol{Y}}(\beta_i)$ as $B \to \infty$ by *Ergodic Theorem*, and (ii) $\frac{1}{B}\sum_{b=1}^{B}\beta_i^{(b)}$ converges in probability to $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$ as $B \to \infty$ by the weak law of large numbers for dependent random sequences. The second result is ensured by the fact that, assuming without loss of generality that true variance of $\beta_i^{(b)}$ under the full conditional density is unity and letting ρ_s denote the correlation between $\beta_i^{(b)}$ and $\beta_i^{(b+s)}$, $Var(\frac{1}{B}\sum_{b=1}^{B}\beta_i^{(b)}) = \frac{1}{B}\{1 + 2\sum_{s=1}^{B-1}(1 - sB^{-1})\rho_s\} \to 0$ as $B \to \infty$ for a convergent Markov chain. Then the effect of the hyperparameters of the variance priors on the main estimator of interest, namely on $E_{\beta_i|\boldsymbol{Y}}(\beta_i)$, is studied through the effect of those on $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$. In this quest, in a way, $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$ acts as a surrogate for $E_{\beta_i|\boldsymbol{Y}}(\beta_i)$.

Second, $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}$ is reexpressed as a composite function of the hierarchical variance components. This is accomplished by evaluating $E_{\beta_i|\Theta_{-\beta_i},\boldsymbol{Y}}(\beta_i)$ at $\tau_\beta = E_{\tau_\beta|\Theta_{-\tau_\beta},\boldsymbol{Y}}(\tau_\beta)$ and $\tau_c = E_{\tau_c|\Theta_{-\tau_c},\boldsymbol{Y}}(\tau_c)$ for the hyperpriors in the IG, UNF and HC models and at $\theta_2 = E_{\theta_2|\Theta_{-\theta_2},\mathbf{Y}}(\theta_2)$ and $\theta_3 = E_{\theta_3|\Theta_{-\theta_3},\mathbf{Y}}(\theta_3)$ for the hyperpriors in the MVLG, MVN and MVSN models. Lastly, directional derivatives are used to investigate the change in $E_{\beta_i|\Theta_{-\beta_i}}(\beta_i)$ with respect to the changes in hyperparameters of the variance components. Let f stand for $E_{\beta_i|\Theta_{-\beta_i},\mathbf{Y}}(\beta_i)$ and D be a domain in $\mathbb{R}^K$ where f is continuously differentiable. In our context, D is the parameter space consisting of the hyperparameters of the variance components and the other parameters in the model. Let $S^K = \{\vec{u} = (u_1, \ldots, u_K) : \sum u_k^2 = 1\}$ be a unit sphere in $\mathbb{R}^K$. Then consider the function $F : D \times S^K \to \mathbb{R}, F(P, \vec{u}) = D_u f(P)$, where $D_u f(P)$ is the directional derivative of the function f at $P \in D$ in direction $\vec{u}$. We study the behavior of F on a sample set $E = \{(P_g, \vec{u}_g) \in D \times S^K, g = 1, \ldots, G\}$. Cleverly chosen E makes investigation of the sensitivity of $E_{\beta_i|\Theta_{-\beta_i},\mathbf{Y}}(\beta_i)$ to the hyperparameters of variance components practicable despite many model parameters with differing ranges. A wide grid of values are considered for the hyperparameters of the variance components. Other parameters involved in $E_{\beta_i|\Theta_{-\beta_i},\mathbf{Y}}(\beta_i)$ are fixed at realistic values. For all the models, direction vector $\vec{u}_g$ is generated randomly for each P_g. Graphical representations of directional derivatives are given in Figs. 1–3 where the vertical axis is the directional derivative. In these graphs, the direction vectors $\vec{u}_g (g = 1, \ldots, G)$ are fixed at the same direction vector which we denote by $\vec{u}_0$. Graphs constructed based on different randomly generated $\vec{u}_g$'s for each P_g are provided in Appendix C. The figures in Appendix C are pebbly unlike the smooth figures below because different direction vectors are plotted on the same figure. Nevertheless the essence is exactly the same. In each panel below, various diverse values are considered for the parameters given at the bottom of the figure. Same functional shape is observed for all of them. Therefore to save space, we present only the ones seen in the panel. Each figure is a graph of rate of change in $E_{\beta_i|\Theta_{-\beta_i},\mathbf{Y}}(\beta_i)$ in a randomly selected direction $\vec{u}_0$ with respect to the hyperparameters specified on the axes. Each graph should be observed for the following two situations i. whether the function f follows a flat pattern, ii. whether the flatness occurs at 0 on the vertical axis. The Bayesian implication of this is as follows. The full conditional expectation of the parameter is equally sensitive to the hyperparameters over the region on which f lies on a flat surface. Otherwise the degree to which the full conditional expectation is sensitive to a small perturbation on the hyperparameters depends on the specific values of the hyperparameters. If f lies flat on the surface at which the vertical axis is 0, there is no change in $E_{\beta_i|\Theta_{-\beta_i},\mathbf{Y}}(\beta_i)$ with changes in the hyperparameters.

Fig. 1 shows the changes in the full conditional expectation of β_i with respect to the changes in the parameters of the Inverse-Gamma priors induced on the variance components. According to Fig. 1(a), f is equally sensitive over the subregion $(\epsilon_\beta, \eta_\beta) \in (c, \infty) \times \mathbb{R}$ where c seems to be somewhere around 50. In addition, f is less sensitive to the hyperparameters on the subregion $(\epsilon_\beta, \eta_\beta) \in (c, \infty) \times \mathbb{R}$ which corresponds to an informative Inverse-Gamma prior for τ_β than on the subregion $(\epsilon_\beta, \eta_\beta) \in (0, c) \times \mathbb{R}$ which corresponds to a noninformative Inverse-Gamma. Other panels of this figure suggest similar conclusion. This finding implies that the users of noninformative variance priors should be cautious in that the posterior output is more sensitive to the choice of the parameters of the noninformative Inverse-Gamma prior for τ_β.

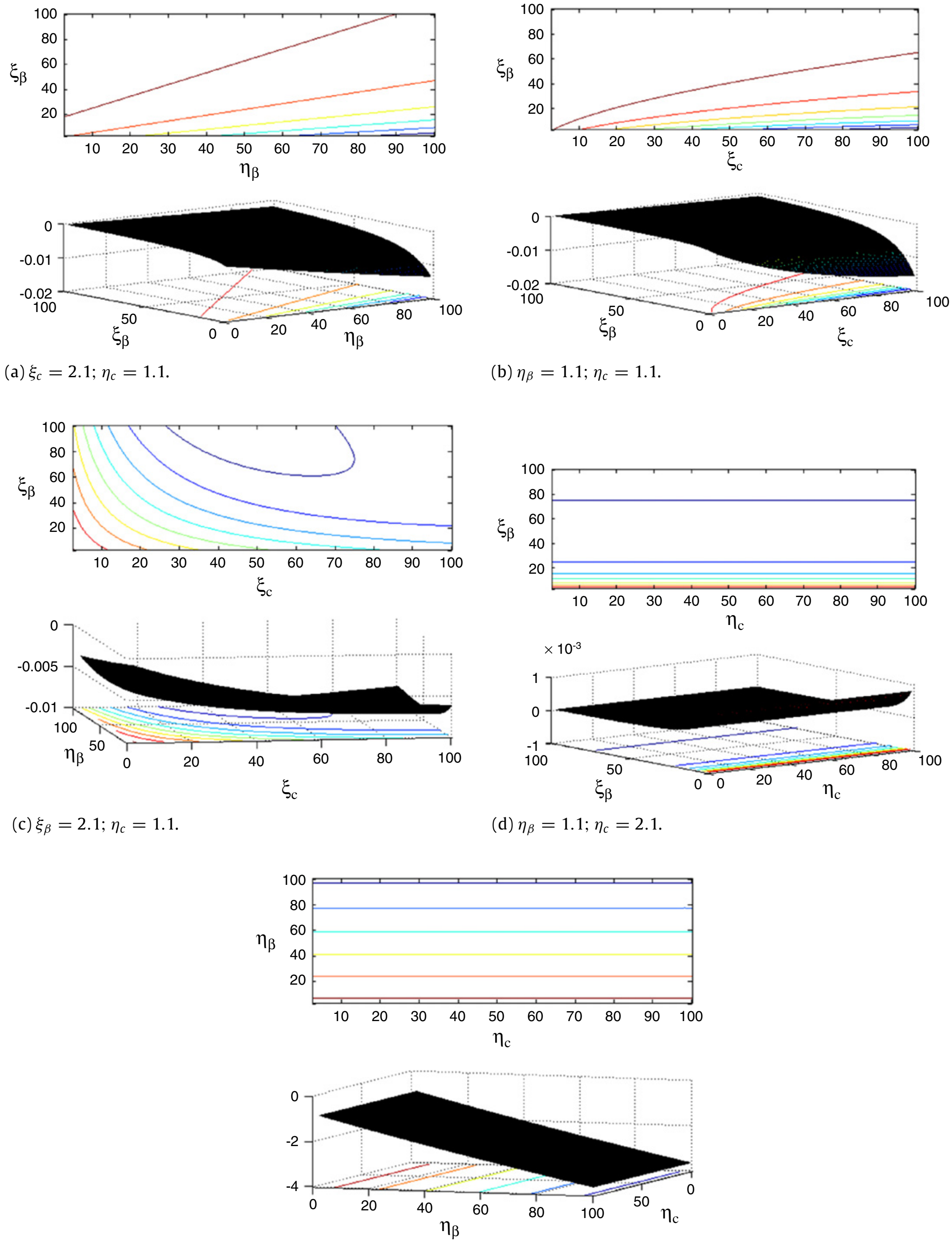
The figures for changes in the full conditional expectation of β_i with respect to the parameters of Uniform prior induced for τ_β and those of Inverse-Gamma induced for τ_c are presented in Fig. 2. Fig. 2(a) indicates that the full conditional expectation of interest is less sensitive to smaller values of b in which the Uniform prior for τ_β is rather informative, when b is considered along with ϵ_c. However, as seen in Fig. 2(b), the result is reversed when b is considered along with η_c. This is due to the interactive role τ_β and τ_c priors play on the full conditional distribution of β_i. This conflicting feature should prevent utilizing Uniform distribution as reference prior for variance components in a multilevel model.

Fig. 3 presents the changes in the full conditional expectation of β_i with respect to the changes in the parameters of the G-MVLG prior induced on the variance components. On account of panels (a), (b), and (c) of Fig. 3, full conditional expectation of β_i has the same degree of sensitivity across the space except when δ and $\eta_.$ are somehow large which is related with an informative setting. Unlike the preceding priors, it can firmly be concluded that degree of sensitivity of the posterior inference is constant over the noninformative hyperparameter space.

In summary, while the changes in the full conditional expectation of β_i with respect to the changes in the parameters of the G-MVLG prior are constant in the region that gives noninformative setting, those with respect to the parameters of univariate models IG and UNF constantly change in the regions giving noninformative setting.

4. Application

In order to illustrate our modeling approaches over a real data set, we revisit a data set used by Moesteller and Tukey [19, p. 503] which is republished by Hand et al. [12, p. 403]. Specific heats of water were measured by six experimenters ($n = 6$) at temperatures 5, 10, 15, 20, 25, and 30 resulting in $J = 6$. Our interest is to fit the model given in (2) using the prior settings discussed in Section 2. For this model, $x_1 = 5, x_2 = 10, x_3 = 15, x_4 = 20, x_5 = 25, x_6 = 30$. Means, standard deviations, lower and upper bounds of 95% highest probability density (HPD) intervals of the posterior distributions of the parameters, and $\hat{R}$ values for the considered models are given in Appendix D. Deviance information criteria (DICs) are given in Table 4. The most striking result is the difference in the posterior inference on the variance components between the two sets of models (MVLG1,MVLG2,MVN1,MVN2,MVNS1,MVNS2) and (HC,UNF,IG). While the point estimates of the variance components vary near 1 for the first set of models, all of them inflate tremendously when IG model is used and for HC and UNF models stage 1 variance inflates. Point estimates of the random effects β_is obtained from these models are quite



(a) $\xi_c = 2.1$; $\eta_c = 1.1$.

(b) $\eta_\beta = 1.1$; $\eta_c = 1.1$.

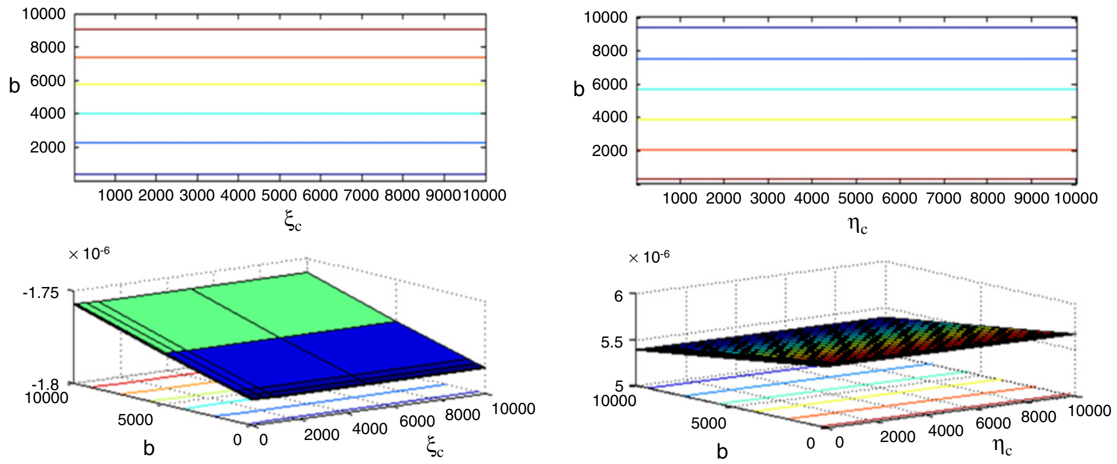
(c) $\xi_\beta = 2.1$; $\eta_c = 1.1$.

(d) $\eta_\beta = 1.1$; $\eta_c = 2.1$.

(e) $\xi_\beta = 2.1$; $\xi_c = 2.1$.

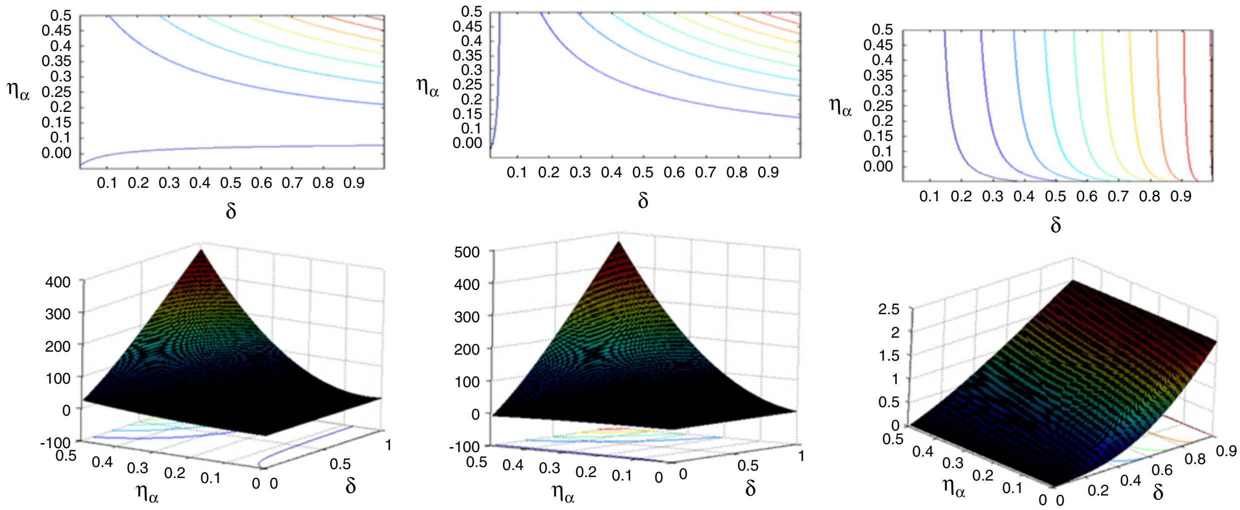
Fig. 1. Changes in the full conditional expectation of β_i for the IG model.

similar. However the posterior standard deviations and the HPD intervals that are used to make inferences about β_is differ depending on the choice of joint or univariate modeling of variance components. For instance, the lengths of HPD intervals for β_is vary between 0.27 and 0.29 under MVLG1 model while it is between 0.29 and 0.52 under HC model. The MVLG2 model produces the best result according to the DIC. It is followed by MVLG1, MVN2, MVN1, MVSN2, and MVSN1. The DICs under the univariate models are at least double those under multivariate models for variance components. This example clearly illustrates that accounting for the intrinsic link among the stage 1 and stage 2 variance components in Bayesian analysis of random effect models yields a better fit.



(a) $\eta_c = 1.1$. (b) $\xi_c = 2.1$.

Fig. 2. Changes in the full conditional expectation of β_i for the UNF model.



(a) $v = 0.5$, $\eta_\beta = 0.1$; $\eta_c = 0.1$. (b) $v = 0.5$, $\eta_\alpha = 0.1$; $\eta_c = 0.1$. (c) $v = 0.5$, $\eta_\alpha = 0.1$; $\eta_\beta = 0.1$.

Fig. 3. Changes in the full conditional expectation of β_i for the MVLG model.

Table 4
DIC values for the fitted models.

Model	DIC
HC	196.279
Unif	199.374
IG	213.229
MVLG	94.392
MVN	95.008
MVSN	97.537
MVLG2	94.131
MVN2	94.972
MVSN2	97.488

5. Discussion

We have shown in a hierarchical model that modeling variance components of different stages jointly a priori resulted in better estimation of the parameters. We have also attempted to evaluate the sensitivity of posterior estimates of the random coefficients to the hyperparameters of the variance priors by assessing the sensitivity of their full conditional expectations to those. Our results indicate that generalized multivariate log gamma distribution as the joint prior distribution of the variance components lead to a rather insensitive posterior outcome.

We focused on a random coefficient normal model where the random coefficients have a diagonal covariance matrix. The considerations here can be extended to models with correlated random effects. Kass and Natarajan [13] suggested an inverse Wishart prior for the random effect covariance matrix which reduces to inverse gamma when the dimension of the random effects is equal to 1. However considerations involving inverse Wishart may not be attractive in the multivariate modeling of variance components across the stages of a hierarchical model due to its relation with inverse gamma prior which is now known to be avoided in the analysis of hierarchical models. Alternative considerations may involve stacking all the variance and covariance components of the random effect model along with stage 1 variance into a vector and consider G-MVLG prior restricted by the positive definiteness of the random effect covariance matrix. In this article, we worked on a model representing a balanced study design where $j = 1, \ldots, J$. Whether the considerations and results attained in this article can be adopted for unbalanced designs requires further investigation.

Appendix A. Derivations for the MVLG model

Joint posterior distribution of all parameters and hyperparameters in MVLG model is as the following:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \alpha_c, \beta_c, \tau_\alpha, \tau_\beta, \tau_c|\boldsymbol{y}) \propto \prod_{i=1}^{n}\prod_{j=1}^{J} \tau_c^{-0.5} \exp\left\{-0.5\tau_c^{-1}[y_{ij} - (\alpha_i + \beta_i(x_j - \bar{x}))]^2\right\}$$

$$\times \prod_{i=1}^{n} \tau_\alpha^{-0.5}\tau_\beta^{-0.5} \exp\{-0.5[\tau_\alpha^{-1}(\alpha_i - \alpha_c)^2 + \tau_\beta^{-1}(\beta_i - \beta_c)^2]\} p(\alpha_c)p(\beta_c)p(\tau_\alpha, \tau_\beta, \tau_c),$$

$$\boldsymbol{\alpha}, \boldsymbol{\beta}, \alpha_c, \beta_c \in \mathbb{R};\ \tau_\alpha, \tau_\beta, \tau_c \in [0, \infty).$$

When the log transformation is applied it is obtained that $\theta_1 = \log(\tau_\alpha)$, $\theta_2 = \log(\tau_\beta)$ and $\theta_3 = \log(\tau_c)$, and the Jacobian determinant is $|J| = \exp\{\theta_1 + \theta_2 + \theta_3\}$. Then full conditional posterior distribution of $\theta_1, \theta_2, \theta_3$ is obtained as follows:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \alpha_c, \beta_c, \theta_1, \theta_2, \theta_3|\boldsymbol{y}) \propto \prod_{i=1}^{n}\prod_{j=1}^{J} \exp\left\{-0.5\theta_3 - 0.5e^{-\theta_3}[y_{ij} - (\alpha_i + \beta_i(x_j - \bar{x}))]^2\right\}$$

$$\times \prod_{i=1}^{n} \exp\{-0.5(\theta_1 + \theta_2) - 0.5[e^{-\theta_1}(\alpha_i - \alpha_c)^2 + e^{-\theta_2}(\beta_i - \beta_c)^2]\}$$

$$\times \delta^\nu \sum_{k=0}^{\infty} \frac{(1-\delta)^k \prod_{j=1}^{3} \eta_j \lambda_j^{-\nu-k}}{[\Gamma(\nu+k)]^2 \Gamma(\nu) k!} \exp\left\{(\nu + k)\sum_{j=1}^{3} \eta_j\theta_j - \sum_{j=1}^{3} \frac{1}{\lambda_j} \exp\{\eta_j\theta_j\}\right\}$$

$$\propto \exp\left\{-(0.5nJ - 1)\theta_3 - 0.5e^{-\theta_3}S_4\right\} \exp\{-(0.5n - 1)(\theta_1 + \theta_2) - 0.5[e^{-\theta_1}S_\alpha + e^{-\theta_2}S_\beta]\}$$

$$\times \exp\left\{\nu\sum_{j=1}^{3} \eta_j\theta_j - \sum_{j=1}^{3} \frac{1}{\lambda_j} \exp\{\eta_j\theta_j\}\right\} \sum_{k=0}^{\infty} \frac{(1-\delta)^k \prod_{j=1}^{3} \lambda_j^{-k}}{[\Gamma(\nu+k)]^2 k!} \exp\left\{k\sum_{j=1}^{3} \eta_j\theta_j\right\}.$$

In order to obtain a familiar Taylor expansion we use the approximation $\Gamma(\nu + n) \approx \exp(-7.24663 + 2.07728n + 1.9922\nu_i)$ proposed by Demirhan and Hamurkaroglu [7]. Accuracy of this approximation is demonstrated, and comparison of its performance with existing approximations is given by Demirhan and Hamurkaroglu [7]. When the approximation is applied the following is obtained:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \alpha_c, \beta_c, \theta_1, \theta_2, \theta_3|\boldsymbol{y}) \propto \exp\left\{-(0.5nJ - 1)\theta_3 - 0.5e^{-\theta_3}S_4\right\}$$

$$\times \exp\{-(0.5n - 1)(\theta_1 + \theta_2) - 0.5[e^{-\theta_1}S_\alpha + e^{-\theta_2}S_\beta]\} \exp\left\{\nu\sum_{j=1}^{3} \eta_j\theta_j - \sum_{j=1}^{3} \frac{1}{\lambda_j} \exp\{\eta_j\theta_j\}\right\}$$

$$\times \exp\left\{(1-\delta)\prod_{j=1}^{3} \lambda_j^{-1} \exp\left\{-2 \cdot 2.07728 + \sum_{j=1}^{3} \eta_j\theta_j\right\}\right\}$$

$$\propto \exp\left\{\left[-(0.5n - 1) + \nu\eta_1\right]\theta_1 + \left[-(0.5n - 1) + \nu\eta_2\right]\theta_2 + \left[-(0.5nJ - 1) + \nu\eta_3\right]\theta_3\right\}$$

$$\times \exp\left\{S_5 \exp\left\{\eta_1\theta_1 + \eta_2\theta_2 + \eta_3\theta_3\right\} - \lambda_1^{-1}e^{\eta_1\theta_1} - \lambda_2^{-1}e^{\eta_2\theta_2} - \lambda_3^{-1}e^{\eta_3\theta_3}\right\}$$

$$\times \exp\left\{-0.5S_\alpha e^{-\theta_\alpha} - 0.5S_\beta e^{-\theta_\beta} - 0.5S_4 e^{-\theta_c}\right\}.$$

Appendix B. Simulation results

Tables of MSEs for all simulation scenarios are presented in the Supplementary Material that can be found online at http://dx.doi.org/10.1016/j.jmva.2014.12.013.

Appendix C. Change graphs for random directions

Graphs of changes in the full conditional expectation of β_i for the randomly selected directions and MVLG models are presented in the Supplementary Material that can be found online at http://dx.doi.org/10.1016/j.jmva.2014.12.013.

Appendix D. Posterior estimates under considered models in data application

Posterior estimates of parameters of the considered models are presented in the Supplementary Material that can be found online at http://dx.doi.org/10.1016/j.jmva.2014.12.013.

Appendix E. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2014.12.013.

References

[1] Y. Aoki, K. Kasai, H. Yamasue, Age-related change in brain metabolite abnormalities in autism: a meta-analysis of proton magnetic resonance spectroscopy stu, Transl. Psychiatr. 2 (2012) e69.
[2] A. Azzalini, The skew-normal distribution and related multivariate families, Scand. J. Stat. 32 (2005) 159–188.
[3] A. Azzalini, A.D. Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[4] W. Browne, D. Draper, A comparison of Bayesian and likelihood-based methods for fitting multilevel models, Bayesian Anal. 1 (2006) 473–514.
[5] D. Chen, F. Liu, C. Yang, X. Liang, Q. Shang, W. He, Z. Wang, Association between the tph1 a218c polymorphism and risk of mood disorders and alcohol dependence: evidence from the current studies, J. Affect. Disord. 138 (2012) 27–33.
[6] S. Chib, E. Greenberg, Understanding the Metropolis–Hastings algorithm, Amer. Statist. 49 (1995) 327–335.
[7] H. Demirhan, C. Hamurkaroglu, On a multivariate log-gamma distribution and the use of the distribution in the Bayesian analysis, J. Statist. Plann. Inference 141 (2011) 1141–1152.
[8] V.A. de Silva, R. Hanwella, Efficacy and tolerability of venlafaxine versus specific serotonin reuptake inhibitors in treatment of major depressive disorder: a meta-analysis of published studies, Int. Clin. Psychopharmacol. 27 (2012) 8–16.
[9] G. Garia, V. Grosbois, A. Waret-Szkutae, S. Babiuk, P. Jacquiet, F. Roger, Lumpy skin disease in ethiopia: seroprevalence study across different agro-climate zones, Acta Trop. 123 (2012) 101–106.
[10] A. Gelman, Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, 1996, Chapter Inference and monitoring convergence.
[11] A. Gelman, Prior distributions for variance parameters in hierarchical models, Bayesian Anal. 1 (2006) 515–553.
[12] D. Hand, F. Daly, A. Lunn, K. McConway, E. Ostrowski, A Handbook of Small Data Sets, Chapman and Hall, London, 1994.
[13] R. Kass, R. Natarajan, A default conjugate prior for variance components in generalized linear mixed models, Bayesian Anal. 1 (2006) 535–542. comment on article by Browne and Draper.
[14] K. Kovacs, An empirical examination of the location and timing of non-renewals in a farmland differential assessment program, Ann. Reg. Sci. 50 (2013) 245–263.
[15] P. Lambert, A. Sutton, P. Burton, K. Abrams, A. Jones, How vague is vague? a simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS, Stat. Med. 24 (2005) 2401–2428.
[16] S. Mahmood, I. Booker, J. Huang, C. Coleman, Effect of topiramate on weight gain in patients receiving atypical antipsychotic agents, J. Clin. Psychopharmacol. 33 (2013) 90–94.
[17] N. Maneeton, B. Maneeton, M. Srisurapanont, S. Martin, Quetiapine monotherapy in acute phase for major depressive disorder: a meta-analysis of randomized, placebo-controlled trials, BMC Psychiatr. 12 (2012) 160.
[18] A. Menegaki, Growth and renewable energy in europe: a random effect model with evidence for neutrality hypothesis, Energy Econ. 33 (2011) 257–263.
[19] F. Moesteller, J. Tukey, Data Analysis and Regression, Addison-Wesley, Massachusetts, 1977.
[20] R. Natarajan, R. Kass, Reference Bayesian methods for generalized linear mixed models, J. Amer. Statist. Assoc. 95 (2000) 227–237.
[21] R. Natarajan, C. McCulloch, Gibbs sampling with diffuse proper priors: a valid approach to data-driven inference?, J. Comput. Graph. Statist. 7 (1998) 267–277.
[22] N. Polson, J. Scott, On the half-cauchy prior for a global scale parameter, Bayesian Anal. 7 (2012) 887–902.
[23] B. Travassos, D. Arajo, K. Davids, K. OHara, J. Leitão, A. Cortinhas, Expertise effects on decision-making in sport are constrained by requisite response behaviours e a meta-analysis, Psychol. Sport Exerc. 14 (2013) 211–219.
[24] R. Yu, M. Abdel-Aty, M. Ahmed, Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors, Accid. Anal. Prev. 50 (2013) 371–376.