Simultaneous Fixed and Random Effects Selection in Order-restricted Mixed Effects Models

HAIFANG SHI, JIAJIA GE
School of Science,
Civil Aviation University of China,
Tianjin,
CHINA

Abstract: - When dealing with longitudinal data, if we directly select a specific model for modeling without any prior information about the existence of significant random effects before utilizing the mixed model, it may result in the misuse of the model, thereby affecting the final estimation results. This paper investigates a variable selection method that can jointly select both fixed and random effects in Bayesian mixed model under order constraints. This method can effectively prevent model misuse. A computationally feasible Gibbs algorithm is proposed for posterior inference. The performance of our proposal is evaluated by simulated data and two real applications related to Blood lead levels and Ramus bone heights. Results show that the proposed approaches perform very well in various situations.

Key-Words: - Random subject effect, Longitudinal data, Order restriction, Ramus bone heights, Model selection, Two-way ANOVA, Blood lead levels.

Received: April 13, 2024. Revised: September 8, 2024. Accepted: September 29, 2024. Published: October 25, 2024.

1 Introduction

In many applications, researchers have prior knowledge about the underlying parameters that satisfy an order restriction before the data are collected. For example, the researchers measured Ramus bone heights of 20 boys at four time points over 18 months, a natural assumption is that the means of ramus bone sizes (μ_i') satisfy the simple order $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_3$. [1], proposed a one-way ANOVA model with order constraints for this data and find evidence that there are only two growth spurts during the 18 months. However, [2] and [3] argued that random subject effect cannot be easily excluded from the model, especially when time is an explanatory variable. They developed a Bayesian hierarchical mixed model for multiple comparisons of fixed effects with a simple order restriction using mixtures of an exponential distribution and a discrete distribution. This matter raises a challenging problem of performing joint fixed and random effects selection in mixed models under order restriction. Model selection in mixed models without constraints has received substantial interest in recent years. Based on a penalized adaptive likelihood, [4] developed Bayesian variable selection by allowing fixed effects or standard deviations of random effects to be exactly zero in linear mixed models. [5], proposed a nested EM algorithm for variable selection of linear mixed effects. [6], proposed a simple iterative penalized procedure that is capable of simultaneously selecting and estimating both fixed effects and random effects in linear mixed-effects models. [7], integrated the penalized quasi-likelihood estimation framework with a penalization approach that enables simultaneous estimation of model parameters while automatically selecting important variables by imposing sparsity constraints on the coefficients of both fixed effects and random effects. [8], reviewed the methods for variable selection in linear mixed-effects models proposed in recent literature and compared the strengths and weaknesses of various approaches through extensive simulations.

However, to our knowledge, there is little literature on the model selection of fixed effects together with the random effects in the mixed model with order restriction. [2], [3] proposed a Bayesian hierarchical mixed model for repeated measures data with missing values and a simple order restriction, but they did not consider the selection of random effects in the model. This paper develops a novel Bayesian variable selection approach for two-way ANOVA mixed model accounting for order

restrictions. Compared with existing literature, our greatest contribution is that our method can determine whether the fixed effects and random effects in the mixed model are significant, and simultaneously estimate the significant effects. In practical data modeling, when significant prior information is lacking, this can prevent misuse of the model and thereby improve the estimation accuracy of the model. A simple and efficient Gibbs sampler is proposed for posterior inference. The paper is organized as follows. Within Section 2, we describe the vectorized form of the two-way ANOVA mixed model under a simple order. We also propose variable selection procedures for both fixed and random eff ects in the model. Section 3 develops computational strategies for posterior inference. Section 4 conducts simulation studies to evaluate the performance of the proposed method. An analysis of two real applications is presented in Section 5. Section 6 concludes the paper.

2 Model Description

Suppose there are n subjects under investigation, and there are k treatment for each subject. Let y_{ij} be the observation of the response variable, the two-way ANOVA mixed model is then expressed as follows:

$$y_{ij} = \mu_i + b_j + \varepsilon_{ij}, i = 1, \dots, k, j = 1, \dots, n \quad (1)$$

where μ_i is the fixed treatment eff ect (mean) for the i-th treatment, b_j is a random subject eff ect which is $N(0, \sigma_\tau^2)$ random variable, and ε_{ij} is measurement error which is $N(0, \sigma^2)$ random variable. Suppose that the random subject eff ects and the measurement errors are all independent. In practical applications, there are generally the following three types of order constraints.

(i) The simple order $\mu_1 \leq \cdots \leq \mu_k$

(ii) The simple tree order $\mu_1 \geq \mu_i, i = 2, \dots k$

(iii) The umbrella order $\mu_1 \leq \cdots \leq \mu_g \geq \mu_{g+1} \geq \cdots \geq \mu_k$.

For the simple order $\mu_1 \leq \cdots \leq \mu_k$, let $\delta_{m-1} = \mu_m - \mu_{m-1} (2 \leq m \leq k)$. Thus we will have $\mu_m = \mu_1 + \delta_1 + \cdots + \delta_{m-1}$ .Let α_j be a standard normal variable, the vectorized form of the model can be written as:

$$y_j = 1_k \mu_1 + \sum_{i=1}^{k-1} x_i \delta_i + 1_k \sigma_T \alpha_j + \varepsilon_j,$$
$$\delta_1 \geq 0, \dots, \delta_{k-1} \geq 0, j = 1, \dots, n \quad (2)$$

where $y_j = \left(y_{1j}, \dots, y_{kj}\right)'$, $\varepsilon_j = \left(\varepsilon_{1j}, \dots, \varepsilon_{kj}\right)'$, and $1_k = (1, \dots, 1)'$ which is a k × 1 vector. If the parameter means satisfy the simple tree order or the umbrella-order, we can also obtain a model similar to (2) through transformation. The detailed transformation method can be referred to in reference, [1].

By introducing indicator variables, we adopt the method proposed by [9] to select fixed and random eff ects simultaneously and fit the model. Bayesian variable selection received large attention in recent years, a nice review can be found in [10]. Setting $\delta_i = \gamma_i \beta_i$ and $\delta_\tau = \gamma_0 \omega_\tau$, we can rewrite model (2) as:

$$y_j = 1_k \mu_1 + \sum_{i=1}^{k-1} + x_i y_i \beta_i + 1_k \gamma_0 \omega_T \alpha_j + \varepsilon_{j,}$$
$$\beta_1 \geq 0, \dots, \beta_{k-1} \geq 0, j = 1, \dots n, \quad (3)$$

where $\gamma_0, \dots, \gamma_{k-1}$ are binary indicator variables (0 or 1) signifying which predictors are active in the model. The indicator variables are assumed independent Bernoulli prior distributions:

$$\gamma_i \sim Bernoulli(\pi_i), i = 0, \dots, k-1. \quad (4)$$

Following [11], we use a weak prior for π_i, i.e. *Uniform*(0,1).We further assign the following hierarchical prior distribution for β_i,

$$\beta_i \sim TN(0, \eta_i, 0, +\infty), \eta_i \sim IGamma(a_1, b_1)$$

where a_1, b_1 are constants, $TN(\mu, \sigma^2, a, b)$ denotes a truncated normal distribution on the interval (a,b), *IGamma*(a, b) denotes an inverse-gamma distribution with density function:

$$f(x|a,b) = \frac{b^a}{T(a)} x^{-a-1} exP^{-\frac{b}{x}}, \quad x > 0.$$

To implement the Bayesian model, we further set a conjugate norm distribution $N(\mu_0, \tau_0^2)$ for μ_1, where τ_0^2 is a constant, and a noninformative joint prior for ω_τ^2 and σ^2,

$$\omega_\tau^2, \sigma^2 \sim \frac{1}{\sigma^2 (k\gamma_0 \omega_\tau^2 + \sigma^2)}$$

3 Posterior

We outline the Gibbs sampler used to obtain posterior samples. The detailed algorithm is as follows.

Let $y=(y_1,\dots,y_n)$,by integrating out b_i ,the marginal likelihood is:

$$\mathrm{m}(\mathrm{y}|\mu_1,\{\beta_i\},\omega_\tau^2,\sigma^2,\{\gamma_i\}) \propto (\sigma^2)^{-\frac{n(k-1)}{2}}(k\gamma_0\omega_\tau^2+\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(s_1-\frac{\gamma_0\omega_\tau^2 S_2}{k\gamma_0\omega_\tau^2+\sigma^2}\right)\right.$$

where

$$s_1=\sum_{j=1}^{n}\sum_{i=1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)^2$$

and

$$s_2=\sum_{j=1}^{n}\left[\sum_{i=1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)\right]^2$$

3.1.1 Step 1–Sampling σ^2 and σ_τ^2:

For the convenience of implementation, we let $\tau^2=k\gamma_0\omega_\tau^2+\sigma^2$.

Update σ^2 from its conditional distribution, an inverse-gamma distribution:

$$\sigma^2|y,\mu_1,\{\gamma_i\},\{\beta_i\},\tau^2 \sim IGamma\left(\frac{n(k-1)}{2},\frac{1}{2}\left(s_1-\frac{s_2}{k}\right)\right.$$

Update τ^2 from its conditional distribution, an inverse-gamma distribution:

$$\tau^2|y,\mu_1,\{\gamma_i\},\{\beta_i\},\sigma^2 \sim IGamma\left(\frac{n}{2},\frac{s_2}{2k}\right)$$

The variance of random subject eff ect σ_τ^2 can be computed by $\frac{\tau^2-\sigma^2}{k}$ after knowing τ^2 and σ^2 .

3.1.2 Step 2–Sampling $\boldsymbol{\beta_p}$:

Update β_p from its conditional distribution, a truncated norm distribution:

Setting

$$s_1=\sum_{j=1}^{n}\sum_{i=1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)^2 \propto \sum_{j=1}^{n}\sum_{i=p+1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)^2 = \left(n\sum_{i=p+1}^{k}x_{ip}^2\gamma_p\right)\beta_p^2-2\beta_p \sum_{j=1}^{n}\sum_{i=p+1}^{k}\left[x_{ip}\gamma_p\left(y_{ij}-\mu_1-\sum_{\substack{l=1\\l\neq p}}^{k-1}x_{il}\gamma_l\beta_l\right)\right]$$

and

$$s_2=\sum_{j=1}^{n}\left[\sum_{i=1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)\right]^2 \propto \sum_{j=1}^{n}\left[\sum_{i=p+1}^{k}\left(y_{ij}-\mu_1-\sum_{l=1}^{k-1}x_{il}\gamma_l\,\beta_l\right)\right]^2 = n\left(\sum_{i=p+1}^{k}x_{ip}\gamma_p\right)^2\beta_p^2-2\beta_p \sum_{j=1}^{n}\sum_{i=p+1}^{k}\left[\left(\sum_{i=p+1}^{k}x_{ip}\gamma_p\right)\left(y_{ij}-\mu_1-\sum_{\substack{l=1\\l\neq p}}^{k-1}x_{il}\gamma_l\beta_l\right)\right] .$$

Setting

$$s_4=\sum_{j=1}^{n}\sum_{i=p+1}^{k}\left[x_{ip}\gamma_p\left(y_{ij}-\mu_1-\sum_{\substack{l=1\\l\neq p}}^{k-1}x_{il}\gamma_l\beta_l\right)\right],$$

and

$$s_5=\sum_{j=1}^{n}\sum_{i=p+1}^{k}\left[\left(\sum_{i=p+1}^{k}x_{ip}\gamma_p\right)\left(y_{ij}-\mu_1-\sum_{\substack{l=1\\l\neq p}}^{k-1}x_{il}\gamma_l\beta_l\right)\right],$$

we then have:

$$\left[\beta_p|y,\mu_1,\omega_\tau^2,\sigma^2,\beta_{-p},\{\gamma_i\}\right] \propto exp\left\{-\frac{1}{2\sigma^2}\left[\left(\sum_{i=p+1}^{k}x_{ip}^2\gamma_p-\frac{\gamma_0\omega_\tau^2}{\tau^2}\left(\sum_{i=p+1}^{k}x_{ip}\gamma_p\right)^2\right)n\beta_p^2-2\beta_p\left(s_4-\frac{\gamma_0\omega_\tau^2}{\tau^2}s_5\right)\right]\right\} exp\left\{-\frac{\beta_p^2}{2n_p}\right\} \quad (5)$$

where $\beta_{-p}=(\beta_1,\dots,\beta_{p-1},\beta_{p+1},\dots,\beta_{k-1})$.

Then the full conditional posterior distributions of β_p is a truncated norm distribution,

$$\beta_p|y,\mu_1,\omega_\tau^2,\sigma^2,\beta_{-p},\{\gamma_i\} \sim \mathrm{TN}\left(\frac{g_2}{g_1},\frac{1}{g_1},0,+\infty\right)$$

where $g_1=\frac{n}{\sigma^2}\left[\sum_{i=p+1}^{k}x_{ip}^2\gamma_p-\frac{\gamma_0\omega_\tau^2}{\tau^2}\left(\sum_{i=p+1}^{k}x_{ip}\gamma_p\right)^2\right]+\frac{1}{n_p}$ and $g_2=\frac{1}{\sigma^2}\left(s_4-\frac{\sigma_\tau^2}{\tau^2}s_5\right)$.

3.1.3 Step 3–Sampling $\boldsymbol{\mu_1}$:

Update μ_1 from its conditional distribution, a norm distribution:

$$[\mu_1|y,\{\beta_i\},\sigma_\tau^2,\sigma^2,\{\gamma_i\}] \propto exp\left\{-\frac{1}{2\sigma^2}\left(s_1-\frac{\gamma_0\omega_\tau^2 S_2}{k\gamma_0\omega_\tau^2+\sigma^2}\right)\right\} exp\left\{-\frac{(\mu_1-\mu_0)^2}{2\tau_0^2}\right\}exp\left\{-\frac{(\mu_1-\mu_0)^2}{2\tau_0^2}\right\} = \exp\left\{-\frac{1}{2}\left[\mu_1^2\left(\frac{1}{\tau_0^2}+\frac{nk}{k\gamma_0\omega_\tau^2+\sigma^2}\right)-2\mu_1\left(\frac{\mu_0}{\tau_0^2}+\frac{s_3}{k\gamma_0\omega_\tau^2+\sigma^2}\right)\right]\right\},$$

where $s_3=\sum_{j=1}^{n}\sum_{i=1}^{k}\left(y_{ij}-\sum_{l=1}^{k-1}x_{il}\gamma_l\beta_l\right)$.

Setting $\mu=\frac{\mu_0}{\tau_0^2}+\frac{s_3}{k\gamma_0\omega_\tau^2+\sigma^2}$ and $v=\frac{1}{\tau_0^2}+\frac{nk}{k\gamma_0\omega_\tau^2+\sigma^2}$, the conditional posterior distributions of μ_1 is then a norm distribution:

$$\mu_1|y,\{\beta_i\},\omega_\tau^2,\sigma^2,\{\gamma_i\} \sim N\left(\frac{n}{v},\frac{1}{v}\right)$$

3.1.4 Step 4–Sampling η_i:

Update η_i from its conditional distribution, an inverse-gamma distribution:

$$\eta_i|y,\{\beta_i\},\omega_\tau^2,\sigma^2,\{\gamma_i\},\mu_1 \sim IGamma\left(\frac{1}{2}+a_{0,}\frac{\beta_i^2}{2}+b_0\right),$$

3.1.5 Step 5-Sampling γ_i:

Let $\gamma_{-i}=(\gamma_0,\dots,\gamma_{i-1},\gamma_{i+1},\dots,\gamma_{k-1})$. It can be demonstrated that the indicator variable γ_i follows a Bernoulli with probability parameter:

$$P(\gamma_i=1|y,\{\beta_i\},\omega_\tau^2,\sigma^2,\gamma_{-i})=\frac{C_i}{C_i+d_i}, \quad (6)$$

where

$$c_i=f(y|\{\beta_i\},\omega_\tau^2,\sigma^2,\gamma_i=1,\gamma_{-i})f(\gamma_i=1,\gamma_{-i})$$

and

$$d_i=f(y|\{\beta_i\},\omega_\tau^2,\sigma^2,\gamma_i=0,\gamma_{-i})f(\gamma_i=0,\gamma_{-i})$$

3.1.6 Step 6-Sampling π_i:

Conditional posterior for π_i. By the prior on π_i and the prior on γ_i, the full conditional distribution of π_i is given by:

$$\pi_i|y,\{\beta_i\},\omega_\tau^2,\sigma^2,\{\gamma_i\},\mu_1 \sim Beta(0.5+\gamma_i,0.5-\gamma_i+1).$$

4 Simulation Studies

In this section, we demonstrate the performance of our methods (BMS) and compare it with the Bayesian procedure for order restricted mixed model proposed by [2], [3] using a series of simulations. We first perform simulations with independent data and then those with dependent data.

4.1 Dependent Data

The data are generated from the model given by:

$$y_{ij}=\mu_i+b_j+\varepsilon_{ij}, i=1,\dots,4, j=1,\dots,n,$$

where random subject effect b_j and error term ε_{ij} are generated independently from $N(0,2)$. Under the simple order and k =4 there are 8 candidate models on the equality/inequality of fixed treatment effects:

$H_0:\mu_1=\mu_2=\mu_3=\mu_4$ $H_{12}:\mu_1=\mu_2=\mu_3<\mu_4$
$H_1:\mu_1=\mu_2<\mu_3<\mu_4$ $H_{13}:\mu_1=\mu_2<\mu_3=\mu_4$
$H_2:\mu_1<\mu_2=\mu_3<\mu_4$ $H_{23}:\mu_1<\mu_2=\mu_3=\mu_4$
$H_3:\mu_1<\mu_2<\mu_3=\mu_4$ $H_F:\mu_1<\mu_2<\mu_3<\mu_4$

Following [12], [1] ,there are three scenarios:

- Case 1: μ = (0, 0, 0, 0)′ , that is, there exists equal fixed treatment effects,
- Case 2: μ = (0, 0, 0, 1)′ , that is, the last group has a different fixed treatment effects,
- Case 3: μ = (1, 2, 3, 4)′ , that is, the fixed treatment effects satisfy simple ordering.

4.2 Independent Data

The model is the same as that in dependent data, except for random effects: $b_j=0, j=1\dots,n$. We consider three sample sizes *n*= 10, *n* = 30, *n* = 100 and repeat 500 times in each example. In all simulation settings, we suppose independent flat inverse Gamma prior distribution *IGamma*(2.2, 20) for η_i, i = 1, 2, 3 and η_τ, choose hyper-parameter $\tau_0^2=100$ for μ_1 such that we obtain weakly informative priors. We run our Gibbs sampler for 10000 iterations with 3000 for burn-in. For fixed treatment effects, Table 1-2 list average posterior probabilities of all the possible models and the percentages of selecting the correct fixed parameters from 500 data sets for dependent data and independent data respectively. The true model is marked as '*'.The bold font is used to highlight the posterior probabilities of the true model. Comparing the conclusions from Table 1, when μ = (0, 0, 0, 0), [2], [3] offer larger percent- ages of selecting the correct model than BMS, but when μ = (0, 0, 0, 1) and μ = (1, 2, 3, 4), our proposed methods BMS generally perform better than the [2], [3] method, showing the good performance of the method. It is also clearly seen from Table 1 that BMS tends to provide a larger average posterior probability of the true model than [2], and [3] in nearly all cases across the examples. Furthermore, for independent data, we can observe similar results from Table 2 as independent data. Moreover, as expected, we see that the average posterior probabilities of the correct model and the percentages of selecting the correct model values for both methods increase as the sample size increases, especially in Case 3 where the fixed treatment effects satisfy simple ordering.

To check the performance of the proposed methods in identifying the correct model for random effect, Table 3 summarizes the percentages of selecting the correct random effect parameters from 500 data sets for dependent data and independent data. Overall, Table 3 indicates that BMS yields promising results in both cases, even for a small sample size when η_i = 10, suggesting good performance of our method.

5 Real Data Example

5.1 Treatment of Lead-exposed Children Trial

We first apply the proposed methodology to the Treatment of Lead- exposed Children trial data (TLC), [13]. In this study, Blood lead levels for 50 of the children who did not receive the succimer capsules were measured at week 0 (baseline), week 1, week 4, and week 6. We let μ_1, μ_2, μ_3 and μ_4 denote the mean blood lead levels corresponding to week 6, week 4,week 1, and baseline, respectively. Because the homes of these children were cleaned using an established TLC regimen, it is reasonable to assume that the mean blood lead levels satisfy the simple order restriction, i.e., $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$.So there are eight candidate models:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ $H_{12}: \mu_1 = \mu_2 = \mu_3 < \mu_4$
$H_1: \mu_1 = \mu_2 < \mu_3 < \mu_4$ $H_{13}: \mu_1 = \mu_2 < \mu_3 = \mu_4$
$H_2: \mu_1 < \mu_2 = \mu_3 < \mu_4$ $H_{23}: \mu_1 < \mu_2 = \mu_3 = \mu_4$
$H_3: \mu_1 < \mu_2 < \mu_3 = \mu_4$ $H_F: \mu_1 < \mu_2 < \mu_3 < \mu_4$

Table 1. Results of the average posterior probabilities of all the possible models and the percentages of selecting the correct model from 500 repetitions for dependent data

Hypothesis		$n_i = 10$		$n_i = 30$		$n_i = 100$	
		BMS	Shang	BMS	Shang	BMS	Shang
case 1	H_0^*	0.560	0.328	0.703	0.335	0.779	0.356
	H_1	0.019	0.062	0.009	0.063	0.004	0.060
	H_2	0.025	0.068	0.010	0.067	0.005	0.062
	H_3	0.020	0.075	0.008	0.070	0.004	0.066
	H_{12}	0.127	0.131	0.094	0.138	0.068	0.137
	H_{13}	0.108	0.142	0.081	0.141	0.069	0.14
	H_{23}	0.137	0.158	0.094	0.152	0.070	0.147
	H_F	0.003	0.036	0.001	0.035	0.000	0.033
	Percentages	0.820	0.998	0.924	0.994	0.962	0.994
case 2	H_0	0.233	0.240	0.070	0.134	0.000	0.010
	H_1	0.075	0.100	0.092	0.153	0.083	0.199
	H_2	0.083	0.100	0.095	0.149	0.077	0.204
	H_3	0.025	0.068	0.008	0.040	0.000	0.003
	H_{12}^*	0.370	0.187	0.644	0.297	0.831	0.475
	H_{13}	0.116	0.126	0.056	0.076	0.003	0.006
	H_{23}	0.086	0.121	0.024	0.064	0.000	0.004
	H_F	0.013	0.059	0.010	0.087	0.005	0.101
	Percentages	0.514	0.244	0.874	0.794	0.978	0.994
case 3	H_0	0.004	0.065	0.000	0.004	0.000	0.000
	H_1	0.179	0.152	0.141	0.205	0.009	0.107
	H_2	0.236	0.134	0.166	0.142	0.009	0.05
	H_3	0.155	0.139	0.158	0.164	0.007	0.042
	H_{12}	0.057	0.111	0.001	0.031	0.000	0.000
	H_{13}	0.139	0.162	0.043	0.141	0.000	0.005
	H_{23}	0.056	0.093	0.001	0.022	0.000	0.000
	H_F^*	0.174	0.145	0.490	0.293	0.975	0.796
	Percentages	0.086	0.036	0.582	0.398	0.998	0.962

Table 2. Results of the average posterior probabilities of all the possible models and the percentages of selecting the correct model from 500 repetitions for independent data

Hypothesis		$n_i = 10$		$n_i = 30$		$n_i = 100$	
		BMS	Shang	BMS	Shang	BMS	Shang
case 1	H_0^*	0.561	0.326	0.703	0.35	0.78	0.372
	H_1	0.019	0.062	0.008	0.060	0.004	0.057
	H_2	0.025	0.068	0.010	0.063	0.005	0.058
	H_3	0.020	0.072	0.008	0.063	0.004	0.06
	H_{12}	0.127	0.137	0.093	0.144	0.067	0.142
	H_{13}	0.108	0.144	0.081	0.141	0.069	0.142
	H_{23}	0.137	0.159	0.095	0.149	0.070	0.144
	H_F	0.003	0.032	0.001	0.028	0.000	0.025
	Percentages	0.832	0.906	0.924	0.940	0.962	0.96
case 2	H_0	0.232	0.190	0.071	0.071	0.000	0.001
	H_1	0.074	0.118	0.093	0.175	0.083	0.199
	H_2	0.083	0.115	0.095	0.171	0.077	0.202
	H_3	0.025	0.066	0.008	0.028	0.000	0.001
	H_{12}^*	0.370	0.221	0.643	0.377	0.831	0.509
	H_{13}	0.116	0.121	0.057	0.055	0.003	0.002
	H_{23}	0.087	0.106	0.024	0.038	0.000	0.000
	H_F	0.013	0.063	0.010	0.084	0.005	0.086
	Percentages	0.510	0.386	0.872	0.828	0.978	0.974
case 3	H_0	0.003	0.010	0.000	0.000	0.000	0.000
	H_1	0.180	0.152	0.144	0.129	0.008	0.008
	H_2	0.235	0.183	0.169	0.163	0.009	0.008
	H_3	0.155	0.172	0.157	0.189	0.006	0.011
	H_{12}	0.057	0.052	0.001	0.002	0.000	0.000
	H_{13}	0.142	0.147	0.043	0.056	0.000	0.000
	H_{23}	0.054	0.072	0.001	0.003	0.000	0.000
	H_F^*	0.173	0.212	0.485	0.458	0.976	0.974
	Percentages	0.088	0.096	0.574	0.502	0.998	0.994

Table 3. Results of the percentages of selecting the correct random eff ect parameters from 500 repetitions

		n =10	n =30	n =100
dependent	case 1	0.962	0.988	1.000
	case 2	0.954	0.990	1.000
	case 3	0.940	0.996	1.000
independent	case 1	0.852	0.932	0.994
	case 2	0.874	0.952	0.986
	case 3	0.896	0.952	1.000

We consider the similar prior specifications as in Section 4 and generate 100, 000 samples with an initial burn-in of 20, 000 iterations. The posterior probabilities of the indicator variable for random subject effect is 1.0, indicating that there is a high probability of non-negligible random subject effects in the data. We also compare our method with [2], [3].The posterior probabilities of all the possible models for each method are listed in Table 4.We observe Shang method chooses $\mu_1 < \mu_2 < \mu_3 < \mu_4$, whereas $\mu_1 = \mu_2 < \mu_3 < \mu_4$ is more supported by BMS. Table 5 also summarizes the posterior estimates for the parameters. It can be seen from Table 4 that both methods yield similar posterior point estimates, however ,BMS tends to offer more narrow credible intervals than Shang for most

parameters, except σ_τ^2, showing an improvement over Shang.

5.2 Ramus Bone Heights

In this section, we illustrate the proposed method to the Ramus bone heights data (Ramus), which are given by [14]. In this study, the Ramus bone heights of 20 boys were measured at 8 years, 8 .5 years, 9 years, 9.5 year over an 18 month period. We also let μ_1, μ_2, μ_3 and μ_4 denote the mean ramus bone heights corresponding to the four time points, respectively. [1] has applied a one-way ANOVA model with no random subject effects to analyze this dataset. The results of [1] show that H_1 has the largest posterior probability.

Table 4. Results of the posterior probabilities of all the possible models for blood lead level data

Model	TLC		Ramus	
	BMS	Shang	BMS	Shang
H_0	0.0000	0.0535	0.0000	0.0176
H_1	0.3946	0.1890	0.0100	0.0870
H_2	0.2624	0.1898	0.0155	0.1679
H_3	0.0021	0.0584	0.0330	0.1950
H_{12}	0.2012	0.1846	0.0148	0.0236
H_{13}	0.0049	0.0633	0.0000	0.0709
H_{23}	0.000	0.0478	0.0006	0.0918
H_F	0.1349	0.2136	0.9410	0.3462

Again, we use the same prior specifications as in Section 4 and perform MCMC to obtain 80 000 samplers after the 20000 burn-in. Tables 4 and 5 lists the posterior probabilities of all the possible models and the posterior estimates for the parameters respectively. We can clearly see that both methods select the same model, indicating that there are three growth spurts during the 18-month period, which is different from [1]. Actually, the posterior probability of the indicator variable for random subject effect offered by BMS is 1.0, suggesting that it is inappropriate to ignore random subject effects in the model. [1] used a one-way ANOVA model without random effects to analyze this real data, which could be inappropriate.

Table 5 lists the posterior estimates for the parameters. In summary, both BMS and Shang's methods yield similar posterior estimates and standard deviations for the parameters. However, generally speaking, BMS provides shorter posterior confidence intervals. For example, for the parameter σ^2 in Ramus, BMS gives a 95% posterior confidence interval of 0.57, which is much smaller than the 3.05 provided by Shang's method.

Table 5. Posterior means (mean), variances(SD) and 95% credible intervals (CrI) for blood lead level data

Methods			μ_1	σ_τ^2	σ^2
TLC	BMS	Mean	23.680	25.586	5.578
		Var	0.610	32.285	0.433
		CrI	(22.13,25.21)	(16.65,38.80)	(4.43,7.00)
	Shang	Mean	23.535	25.424	6.217
		Var	1.266	32.463	1.278
		CrI	(21.01,25.46)	(16.47,38.64)	(4.66,8.87)
Ramus	BMS	Mean	48.477	6.861	0.734
		Var	0.386	6.737	0.021
		CrI	(47.23,49.68)	(3.46,13.30)	(0.50,1.07)
	Shang	Mean	48.247	6.678	1.452
		Var	1.215	6.771	0.767
		CrI	(49.94,50.22)	(3.235,13.09)	(0.68,3.73)
			δ_1	δ_2	δ_3
TLC	BMS	Mean	0.278	0.427	1.740
		Var	0.173	0.242	0.253
		CrI	(0.00,1.31)	(0.00,1.49)	(0.71,2.67)
	Shang	Mean	0.546	0.507	1.362
		Var	0.836	0.654	1.289
		CrI	(0.00,3.01)	(0.00,2.64)	(0,3.53)
Ramus	BMS	Mean	0.982	0.947	0.856
		Var	0.085	0.093	0.097
		CrI	(0.40,1.55)	(0.32,1.54)	(0.00,1.42)
	Shang	Mean	1.419	0.875	0.632
		Var	1.321	0.830	0.611
		CrI	(0.00,3.70)	(0.00,2.82)	(0.00,2.43)

6 Discussion and Conclusion

In this article, we develop a simultaneous selection method of fixed and random effects in a Bayesian restricted two-way ANOVA mixed model, which can accommodate some constraints such as simple order, tree order umbrella order, etc. Simulation studies show that the proposed Bayesian variable selection approach works well in the selection of fixed and random effects whether in dependent or independent data. Specifically, in both simulations of dependent and independent data, our method not only successfully identifies the correct fixed effects but also effectively determines the presence of random effects. Furthermore, the accuracy of this identification increases with the number of samples, demonstrating the consistency of our method.

Real data examples indicate that the proposed method is likely to provide more narrow 95% credible intervals than the competing method.

Moreover, we find strong evidence that there exists significant random subject effects in Ramus bone heights data. However, [1] analyzed the dataset by using an unsuitable one-way ANOVA model without random effects and selected a different model. This shows that it is necessary to consider a simultaneous selection of fixed and random effects for longitudinal data under order constraints.

References:

[1] Oh, M. S., Shin, D., A unified Bayesian inference on treatment means with order constraints, *Computational Statistics, and Data Analysis*, Vol.55, 2011, pp. 924-934. https://doi.org/10.1016/j.csda.2010.07.026.

[2] Shang, J., A Bayesian multiple comparison procedure for simple order-restricted mixed models with missing values, *Journal of Data Science*, Vol.9, 2011, pp. 311-330. https://doi.org/10.6339/JDS.201107_09(3).0001.

[3] Shang, J., Cavanaugh, J. E., Wright, F. T., A Bayesian multiple comparison procedure for ordered-restricted mixed models. *International Statistical Review*, Vol 76, 2008, pp. 268-284. https://doi.org/10.1111/j.1751-5823.2008.00051.x.

[4] Bondell, H. D., Krishna, A., Ghosh, S. K., Joint variable selection for fixed and random effects in linear mixed-effects models, *Biometrics*, Vol.66, No.4, 2020, pp. 1069–1077. https://doi.org/10.1111/j.1541-0420.2010.01391.x.

[5] Du, Y., Kahlili, A., Neslehova, J. G., Steele, R. J., Simultaneous fixed and random effects selection in finite mixture of linear mixed-effects models, *The Canadian Journal of Statistics*, Vol.41, 2013, pp. 596-616. https://doi.org/10.1002/cjs.11192.

[6] Peng, H., Lu, Y., Model selection in linear mixed effect models, *Journal of Multivariate Analysis*, Vol.109, 2012, pp. 109-129. https://doi.org/10.1016/j.jmva.2012.02.005.

[7] Hui, F. K. C., Muller, S., Welsh, A. H., Joint selection in mixed models using regularized PQL, *Journal of the American Statistical Association*, Vol.112, 2017, pp. 1323-1333. https://doi.org/10.1080/01621459.2016.1215989.

[8] Buscemi, S., Plaia, A., Model selection in linear mixed-effect models, *Biometrika*, Vol.104, No.4, 2020, pp. 529-575. https://doi.org/10.1007/s10182-019-00359-z.

[9] Kuo, L., Mallick, B., Variable selection for regression models, *Sankhya,B*, Vol.60, 1998, pp. 65-81.

[10] Korobilis, D., Shimizu, K., Bayesian approaches to shrinkage and sparse estimation, *Foundations and Trends® in Econometrics*, Vol.11, No.4, 2022, pp. 230-354. http://dx.doi.org/10.1561/0800000041.

[11] Hara, R. B., Sillanpaa, M. J., Review of bayesian variable selection methods: what, how and which, *Bayesian Analysis*, 4, 2009, pp. 85-118. DOI: 10.1214/09-BA403.

[12] Dunson, D. B., Neelon, B., Bayesian inference on order-constrained parameters in generalized linear models, *Biometrics*, Vol.59, 2003, pp. 286-295. https://doi.org/10.1111/1541-0420.00035.

[13] Fitzmaurice, G., Laird, N. & Ware, J., *Applied Longitudinal Analysis*, New York: John Wiley & Sons, 2004. DOI:10.1002/9781119513469.

[14] Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., Bugs 0.5: examples Vol. 2, *MRC Biostatistics Unit*, 1996, [Online]. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=56b11079d6495501c84932c0a7a372ca6bc370ae. The Last Accessed date is August 14, 1996 (Accessed Date: October 1, 2024).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Jiajia Ge carried out the real data analysis.
- Haifang Shi was in charge of the theoretical design and simulation.

The two authors collaboratively completed the writing of the paper.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The first author is supported by the Research Startup Fund of Civil Aviation University of China(Fund No.: 2013QD06S).

Conflict of Interest

The authors have no conflicts of interest to declare.