BAYESIAN ANALYSIS FOR RANDOM COEFFICIENT REGRESSION MODELS USING NONINFORMATIVE PRIORS

by

Ruoyong Yang and Ming-Hui Chen

Purdue University

Worcester Polytechnic Institute

Technical Report #93-40

Department of Statistics

Purdue University

July 1993

Bayesian Analysis for Random Coefficient Regression Models Using Noninformative Priors *

Ruoyong Yang
Department of Statistics
Purdue University
West Lafayette, IN 47907

Ming-Hui Chen†
Department of Mathematical Sciences
Worcester Polytechnic Institute
Worcester, MA 01609

January 20, 1994

Abstract

We apply Bayesian approach, through noninformative priors, to analyze Random Coefficient Regression (RCR) model. The Fisher information matrix, the Jeffreys prior and reference priors are derived for this model. Then, we prove that the corresponding posteriors are proper when the number of full rank design matrices are greater than or equal to twice the number of regression coefficient parameters plus 1, and that the posterior means for all parameters exist if one more additional full rank design matrix is available. A hybrid Markov chain sampling scheme is developed for computing the Bayesian estimators for parameters of interest.

AMS 1991 subject classifications: 62C10, 62F15, 62J99.
Key words and phrases: Bayesian posterior, Fisher information matrix, Gibbs sampler, Jeffreys prior, reference prior, repeated measures.

*This research was supported by National Science Foundation grant DMS-8923071 at Purdue University.
†This work was done when the second author was a Ph.D. Student in the Department of Statistics of the Purdue University.

1 Introduction

Consider the random coefficient regression (RCR) model

$$\underset{\sim}{y}_i = X_i \underset{\sim}{\beta}_i + \varepsilon_i, \tag{1.1}$$

where $\underset{\sim}{y}_i$ is a $t_i \times 1$ vector of observations. X_i is a $t_i \times p$ constant design matrix, $\underset{\sim}{\beta}_i$ is a $p \times 1$ vector of random coefficients for the ith experimental subject and ε_i is a vector of errors for $i = 1, 2, \cdots, n$.

Furthermore we assume that $(\underset{\sim}{\beta}_i, \varepsilon_i, i = 1, 2, \cdots, n)$ are independent and

$$\underset{\sim}{\beta}_i \sim \mathrm{MVN}(\underset{\sim}{\beta}, \Sigma), \quad \text{and } \varepsilon_i \sim \mathrm{MVN}(\underset{\sim}{0}, \sigma^2 I_i), \tag{1.2}$$

where $\underset{\sim}{\beta}$ is a $p \times 1$ vector of the mean of the $\underset{\sim}{\beta}_i$ vectors and I_i is a $t_i \times t_i$ identity matrix.

The RCR model (1.1) arises from that the probability distributions for responses of different individuals belong to a single family and that the regression parameters vary across individuals because of random-effects. It is widely used in growth-curves, medical studies, repeated-measures, or longitudinal studies. A lot of experiments are performed for the study of the population parameters, $\underset{\sim}{\beta}$, σ^2 and Σ. Unfortunately, the inference concerning $\underset{\sim}{\beta}$ for this model (1.1) is a notoriously difficult problem, especially when the data is unbalanced or incomplete, which is the typical case in clinical trials.

The maximum likelihood (ML) estimate and the restricted maximum likelihood (REML) estimate for the RCR model (1.1) were extensively studied by many authors in literature, although the explicit form of the maximum likelihood estimate for this model is not available. Computation of the ML or REML estimate usually resorts to EM algorithm. The literature includes Laird, Lange and Stram [10], Jennrich [9] and Laird and Ware [11]. Harville [8] reviewed the ML and REML approach with some restrictions on the covariance matrix Σ, in the context of ANOVA model.

Bayesian inference based on a noninformative prior for the population parameters is very attractive when both t_i and n are small, or the data is unbalanced. As stated by Harville [8] (Section 8.3), the ML and REML estimators are simply the mode of the posterior and marginal posterior based on a flat prior. Jeffreys prior would produce a shrinkage estimator which has uniformly smaller MSE than both the ML and REML estimators, although it has a little downward bias.

Thus Jeffreys prior may have some appeal for frequentists who care about MSE but not about small bias.

In the case of one way ANOVA, which is the simplest version of the model (1.1), the reference prior has been studied by Berger and Bernardo [2]. Ye [18] also studied the one way ANOVA model by emphasis on estimation of variance components ratio.

We study the general RCR model (1.1) using Bayesian approach through noninformative priors. The outline of this paper is as follows. Section 2 contains several matrix results and the derivation of the Fisher information matrix of the RCR model (1.1). In Section 3, we derive the Jeffreys prior, reference priors and bounds for these priors. In Bayesian analysis, it is important to know if the posteriors are proper or not when the noninformative priors are used. In Section 4, we study the sufficient conditions for the proper of the posterior distributions corresponding to the Jeffreys prior and reference priors and the existence of the posterior moments. In Section 5, a hybrid Markov chain sampling scheme, which is used for computing posterior expectations, is developed, and an illustrative example is also given. Comments and generalizations are given in Section 6.

2 Preliminaries

2.1 The Likelihood

In the model (1.1), $\underset{\sim}{\beta}_i$'s are random parameters and $\underset{\sim}{\beta}$, Σ and σ^2 are population parameters. Therefore, The likelihood function depends only on $\underset{\sim}{\beta}$, Σ and σ^2.

Since $\underset{\sim}{y}_i = X_i\underset{\sim}{\beta}_i + \varepsilon_i$, $\underset{\sim}{\beta}_i \sim \text{MVN}(\underset{\sim}{\beta}, \Sigma)$, $\varepsilon_i \sim \text{MVN}(\underset{\sim}{0}, \sigma^2 I_i)$, and $\underset{\sim}{\beta}_i$ and ε_i are independent, then

$$\underset{\sim}{y}_i \sim \text{MVN}\left(X_i\underset{\sim}{\beta}, X_i\Sigma X_i^t + \sigma^2 I_i\right). \tag{2.1}$$

Therefore the likelihood function, ignoring the constant, is

$$L(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data) = \left[\prod_{i=1}^{n} \left|X_i\Sigma X_i^t + \sigma^2 I_i\right|^{-\frac{1}{2}}\right] \cdot exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta})^t(X_i\Sigma X_i^t + \sigma^2 I_i)^{-1}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta})\right\}, \tag{2.2}$$

where "*data*" is $(\underset{\sim}{y}_i, X_i), i = 1, 2, \cdots, n$.

2.2 Notations and Several Matrix Results

We will use the following notations throughout this paper. A^t, $|A|$ and $\mathrm{tr}(A)$ denote the transpose, determinant and trace of a square matrix, A, respectively. Denote *vec*() to be the matrix operator which arranges the columns of a matrix into one long column, and *vecp*() to be the matrix operator which arranges the columns of lower left corner of a symmetric matrix into one long column. The Kronecker product of two matrices, A and B, is denoted by $A \otimes B$. $A \geq 0$ means that A is semi-definite positive and $A \geq B$ means $A - B \geq 0$. G denotes a $(p(p+1)/2) \times p^2$ constant matrix $\partial(vecV)/\partial(vecpV)$, where V is a $p \times p$ symmetric matrix. c always denotes some constant independent of parameters, while the value of it may vary from place to place.

We will heavily use the following matrix results in this paper. Results 2.1 and 2.2 can be found in J. R. Magnus and H. Neudecker [12] (pp. 30-31). Result 2.4 follows from Result 2.1. Results 2.5 and 2.6 are given by Wiens [16].

Result 2.1 *For any $p \times p$ matrices A, B, C, and D,*

$$\left(vec(D^t)\right)^t \left(C^t \otimes A\right)(vec(B)) = tr(ABCD). \tag{2.3}$$

Result 2.2

$$vec(ABC) = (C^t \otimes A)vec(B). \tag{2.4}$$

Result 2.3 *Suppose* $\begin{pmatrix} A & B^t \\ B & C \end{pmatrix} \geq 0$, *and* $A > 0$, *then*

$$\begin{vmatrix} A & B^t \\ B & C \end{vmatrix} \leq |A| \cdot |C|. \tag{2.5}$$

Result 2.4 *If for $p \times p$ matrices A and B, $0 \leq A \leq B$, then*

$$A \otimes A \leq A \otimes B \leq B \otimes B. \tag{2.6}$$

If $p \times p$ matrices $A_i \geq 0$, $i = 1, \cdots, n$, then

$$\sum_{i=1}^n A_i \otimes A_i \leq \left(\sum_{i=1}^n A_i\right) \otimes \left(\sum_{i=1}^n A_i\right). \tag{2.7}$$

Result 2.5 *If A is a $p \times p$ matrix, and Σ is a $p \times p$ symmetric matrix, then*

$$\left|G(A \otimes A)G^t\right| = |GG^t| \cdot |A|^{p+1}. \tag{2.8}$$

Result 2.6 *For a $p \times p$ symmetric matrix V,*

$$vec(V) = G^t vecp(V). \tag{2.9}$$

2.3 The Fisher Information Matrix for $\underset{\sim}{\beta}$, Σ and σ^2

The following result was given by Tracy and Jinadasa [15].

Lemma 2.1 *If a $p \times 1$ random vector $\underset{\sim}{y} \sim MVN(\underset{\sim}{\mu}, V)$, then the Fisher information matrix for $\underset{\sim}{\mu}$ and V is*

$$I(\underset{\sim}{\mu}, V) = \begin{pmatrix} V^{-1} & 0 \\ 0 & \frac{1}{2}G(V^{-1} \otimes V^{-1})G^t \end{pmatrix}. \tag{2.10}$$

Using Lemma 2.1, the Fisher information matrix for $\underset{\sim}{\beta}$, Σ and σ^2 thus obtains.

Proposition 2.1 *For a RCR model (1.1) with the likelihood function given in (2.2), the Fisher information matrix for $\underset{\sim}{\beta}$, Σ and σ^2 is*

$$I(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2) = \sum_{i=1}^{n} I_i(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2) \tag{2.11}$$

with

$$I_i(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2) = \begin{pmatrix} I_i(\underset{\sim}{\beta}) & 0 & 0 \\ 0 & \frac{1}{2}I_i(\Sigma) & \frac{1}{2}I_i(\Sigma, \sigma^2) \\ 0 & \frac{1}{2}I_i^t(\Sigma, \sigma^2) & \frac{1}{2}I_i(\sigma^2) \end{pmatrix}, \tag{2.12}$$

where

$$I_i(\underset{\sim}{\beta}) = A_i, \quad I_i(\Sigma) = G(A_i \otimes A_i)G^t, \tag{2.13}$$

$$I_i(\Sigma, \sigma^2) = G\ vec(C_i),\ I_i(\sigma^2) = tr\left((X_i \Sigma X_i^t + \sigma^2 I_i)^{-2}\right), \tag{2.14}$$

$$A_i \stackrel{def}{=} X_i^t(X_i \Sigma X_i^t + \sigma^2 I_i)^{-1} X_i,\ C_i \stackrel{def}{=} X_i^t(X_i \Sigma X_i^t + \sigma^2 I_i)^{-2} X_i. \tag{2.15}$$

Proof: See Appendix A. ■

Consider the following transformation:

$$\begin{cases} \underset{\sim}{\beta} = \underset{\sim}{\beta} \\ \Lambda = \Sigma/\sigma^2 \\ \sigma^2 = \sigma^2, \end{cases} \tag{2.16}$$

which will be used to develop the reference prior of the RCR model (1.1).

Proposition 2.2 *The Fisher information matrix for* $\underset{\sim}{\beta}$, Λ *and* σ^2 *is*

$$I(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2) = \sum_{i=1}^{n} I_i(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2), \tag{2.17}$$

with

$$I_i(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2) = \frac{1}{2}\begin{pmatrix} 2B_i/\sigma^2 & 0 & 0 \\ 0 & G(B_i \otimes B_i)G^t & Gvec(B_i)/\sigma^2 \\ 0 & [Gvec(B_i)/\sigma^2]^t & t_i/\sigma^4 \end{pmatrix}, \tag{2.18}$$

where

$$B_i \overset{def}{=} X_i^t(X_i\Lambda X_i^t + I_i)^{-1}X_i = \sigma^2 A_i. \tag{2.19}$$

Proof: Note that

$$I_i(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2) = \left[\frac{\partial(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2)}{\partial(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2)}\right] I_i(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2) \left[\frac{\partial(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2)}{\partial(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2)}\right]^t, \tag{2.20}$$

where

$$\frac{\partial(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2)}{\partial(\underset{\sim}{\beta},\ \Lambda,\ \sigma^2)} = \begin{pmatrix} I_p & 0 & 0 \\ 0 & \sigma^2 I_{p(p+1)/2} & 0 \\ 0 & [vecp(\Lambda)]^t & 1 \end{pmatrix}. \tag{2.21}$$

Thus Proposition 2.2 follows from Proposition 2.1 and Results 2.1, 2.2 and 2.6. ■

3 The Noninformative Priors

3.1 The Jeffreys Prior

Specification of a prior distribution is an important issue in Bayesian analysis. When subjective prior is not available, use of noninformative priors has an extensive tradition in statistics. A commonly used noninformative prior is Jeffreys prior, along with the "uniform" prior ($\pi_U(\underset{\sim}{\beta}, \Sigma, \sigma^2) = 1$).

Use of π_U is generally very successful, although there are concerns about its lack of invariance to transformation. Also, a number of counterexamples to its use have been encountered (see, e.g., Fraser, Monette and Ng [7], and Ye and Berger [19]). Jeffreys prior exhibits many nice features that make it an attractive prior. One such property is parameterization invariance. In one-dimensional parameter problems, Jeffreys prior is also an optimal noninformative prior in the sense that it maximizes the missing information (see Bernardo [6]).

The Jeffreys prior for the RCR model (1.1) immediately follows from Proposition 2.1. Denote $\pi_J(\underset{\sim}{\beta}, \Sigma, \sigma^2)$ be the Jeffreys prior. Then

$$\pi_J(\underset{\sim}{\beta}, \Sigma, \sigma^2) \propto \left|\sum_{i=1}^{n} I_i(\underset{\sim}{\beta})\right|^{\frac{1}{2}} \cdot \left| \begin{array}{cc} \sum_{i=1}^{n} I_i(\Sigma) & \sum_{i=1}^{n} I_i(\Sigma, \sigma^2) \\ \sum_{i=1}^{n} I_i(\Sigma, \sigma^2) & \sum_{i=1}^{n} I_i(\sigma^2) \end{array} \right|^{\frac{1}{2}}, \tag{3.1}$$

where $I_i(\underset{\sim}{\beta})$, $I_i(\Sigma)$, $I_i(\Sigma, \sigma^2)$, and $I_i(\sigma^2)$ are defined in (2.13) and (2.14). The Jeffreys prior w.r.t. the parameterization ($\underset{\sim}{\beta}$, Λ, σ^2), denoted by $\pi_J(\underset{\sim}{\beta}, \Lambda, \sigma^2)$, is

$$\pi_J(\underset{\sim}{\beta}, \Lambda, \sigma^2) \propto \left|\sum_{i=1}^{n} B_i\right|^{\frac{1}{2}} \cdot \left|G[\sum_{i=1}^{n}(B_i \otimes B_i) - \frac{vec(\sum_{i=1}^{n} B_i)(vec(\sum_{i=1}^{n} B_i))^t}{\sum_{i=1}^{n} t_i}]G^t\right|^{\frac{1}{2}} / \sigma^{p+2}, \tag{3.2}$$

where B_i is given by (2.19).

3.2 Reference Priors

Although Jeffreys prior is invariant under reparameterization and has been proven to be a successful noninformative prior for one-dimensional parameter problems, Jeffreys himself, however, noticed difficulties in multi-dimensional parameter problems, especially when nuisance parameters present. In our problem, we are concerning inference of $\underset{\sim}{\beta}$, Σ or σ^2. This puts us in a typical situation where reference prior has been shown to be a very promising noninformative prior (cf., Berger and Bernardo [1], [2], [3] and [4], Ye and Berger [19]). A review and discussion of the current status of the reference prior can be found in Berger and Bernardo [4].

In the following theorem, the reference priors for $\underset{\sim}{\beta}$, Λ and σ^2 obtain. Note that the reference prior depends on the "group ordering" that is typically a listing of parameters according to perceived "importance."

Theorem 3.1 *For the RCR model (1.1), the reference priors for different groups of ordering of* $(\underset{\sim}{\beta}, \sigma^2, \Lambda)$ *are:*

Group Ordering	Reference Prior
$\{\underset{\sim}{\beta}, \sigma^2, \Lambda\}, \{\sigma^2, \underset{\sim}{\beta}, \Lambda\}$ or $\{\sigma^2, \Lambda, \underset{\sim}{\beta}\}$	$\lvert G \sum_{i=1}^{n}(B_i \otimes B_i)G^t\rvert^{\frac{1}{2}}/\sigma^2$
$\{\underset{\sim}{\beta}, \Lambda, \sigma^2\}, \{\Lambda, \underset{\sim}{\beta}, \sigma^2\}$ or $\{\Lambda, \sigma^2, \underset{\sim}{\beta}\}$	$\left\lvert G[\sum_{i=1}^{n}(B_i \otimes B_i) - \frac{vec(\sum_{i=1}^{n} B_i)(vec(\sum_{i=1}^{n} B_i))^t}{\sum_{i=1}^{n} t_i}]G^t\right\rvert^{\frac{1}{2}}/\sigma^2$

(3.3)

where B_i is given by (2.19).

Proof: See Appendix B. ■

3.3 Bounds for Reference Priors and Jeffreys Prior

The results in this section will be used to prove that the posterior distribution is proper and the posterior means for parameters of interest exist when we use the Jeffreys prior or reference priors.

Proposition 3.1

$$|\sum_{i=1}^{n} A_i| \leq \frac{c}{\sigma^{2p}}, \tag{3.4}$$

$$|\sum_{i=1}^{n} A_i| \leq \frac{c}{\lambda\sigma^{2(p-1)}}, \tag{3.5}$$

where λ is the largest eigenvalue of Σ.

Proof: See Appendix C. ■

By Proposition 3.1, the following result is immediate.

Proposition 3.2

$$\pi_J(\underset{\sim}{\beta}, \Sigma, \sigma^2) \leq \frac{c}{\lambda^{\frac{p}{2}}\sigma^{p^2+p+2}}, \tag{3.6}$$

$$\pi_J(\underset{\sim}{\beta}, \Sigma, \sigma^2) \leq \frac{c}{\lambda^{\frac{p+2}{2}} \sigma^{p^2+p}}, \tag{3.7}$$

The bounds for the reference prior w.r.t. the parameterization $\underset{\sim}{\beta}, \Sigma, \sigma^2$ *are*

$$\pi_R(\underset{\sim}{\beta}, \Sigma, \sigma^2) \leq \frac{c}{\sigma^{p^2+p+2}}, \tag{3.8}$$

$$\pi_R(\underset{\sim}{\beta}, \Sigma, \sigma^2) \leq \frac{c}{\lambda^{\frac{p+1}{2}} \sigma^{p^2+1}}, \tag{3.9}$$

Proof: Using Proposition 3.1, Results 2.4 and 2.5. ■

4 Posterior Analysis of the RCR Model

4.1 The Posterior Density for $\underset{\sim}{\beta}$, Σ and σ^2

Denote $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ to be the joint posterior density for $\underset{\sim}{\beta}$, Σ, and σ^2 with prior $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2)$. Then

$$\begin{aligned} \pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data) \quad \propto \quad & \pi(\underset{\sim}{\beta}, \Sigma, \sigma^2) \cdot \prod_{i=1}^{n} \left| X_i \Sigma X_i^t + \sigma^2 I_i \right|^{-\frac{1}{2}} \\ & \cdot \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta})^t (X_i \Sigma X_i^t + \sigma^2 I_i)^{-1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}) \right\}. \end{aligned} \tag{4.1}$$

Due to the complexity of the Jeffreys prior and reference priors, analytic analysis on the posterior is difficult, for example, the explicit form of the marginal posterior of the parameters is not available. It is not easy to see whether the joint posterior density, $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$, is proper or improper. Especially when we deal with the unbalanced observations, it becomes extremely hard. It seems necessary to pose some conditions on the data structures in order to obtain a proper joint posterior density. For example, we need to know how many subjects should be taken and how many full rank design matrices X_i are needed for obtaining a proper posterior.

By using the upper bound of the Jeffreys prior and reference priors in Proposition 3.2, the posteriors can be proven to be proper. In Sections 4.2 to 4.4, we prove the Jeffreys posterior is proper and the corresponding posterior means of parameters exist under certain conditions. Under similar conditions, using same technique, the posterior derived from the reference priors and "uniform" prior can also be proven to be proper and the corresponding posterior means of parameters to be existent (see Theorems 4.3 and 4.4).

By (3.6) and the fact that $\lambda^{p/2} \geq |\Sigma|^{1/2}$, an upper bound of $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is,

$$\begin{aligned} \pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data) \leq & \frac{c}{|\Sigma|^{\frac{1}{2}} \sigma^{p^2+p+2}} \cdot \prod_{i=1}^{n} \left| X_i \Sigma X_i^t + \sigma^2 I_i \right|^{-\frac{1}{2}} \\ & \cdot \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta})^t (X_i \Sigma X_i^t + \sigma^2 I_i)^{-1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}) \right\}, \end{aligned} \quad (4.2)$$

By denoting the RHS of (4.2) as $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$, it is immediate that if $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is proper, then $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is also proper. However, $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is still quite complicated even though it is much simpler than the original posterior density. In the following subsections, our inference will be based on $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ instead of $\pi(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$.

4.2 Auxiliary Variables $\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \cdots, \underset{\sim}{\beta}_n$

The most troublesome part in $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is Σ and σ^2. Therefore, we introduce auxiliary variables $\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n$ to get a higher dimensional integral and then integrate out Σ and σ^2 first.

Let

$$\begin{aligned} & \pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n, \Sigma, \sigma^2 \mid data) \propto \frac{1}{\sigma^{p^2+p+2}} \cdot \frac{1}{\sigma^{pn}} \cdot \frac{1}{|\Sigma|^{\frac{n+1}{2}}} \\ & \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i) - \frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t \Sigma^{-1} (\underset{\sim}{\beta}_i - \underset{\sim}{\beta}) \right\}. \end{aligned} \quad (4.3)$$

It is easy to prove that

$$\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data) \propto \int \pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n, \Sigma, \sigma^2 \mid data) d\underset{\sim}{\beta}_1 \cdots d\underset{\sim}{\beta}_n. \quad (4.4)$$

For simplicity, we call that $\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n, \Sigma, \sigma^2 \mid data)$ is the joint posterior density for $\underset{\sim}{\beta}$, $\underset{\sim}{\beta}_1, \cdots$, $\underset{\sim}{\beta}_n$, Σ, and σ^2. Then $\pi^*(\underset{\sim}{\beta}, \Sigma, \sigma^2 \mid data)$ is the marginal posterior density for $\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n, \Sigma, \sigma^2 \mid data)$. So it suffices to prove that $\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n, \Sigma, \sigma^2 \mid data)$ is proper.

By integrating out Σ and σ^2, the marginal posterior density for $\underset{\sim}{\beta}$, $\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n$ is

$$\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n \mid data) \propto \int \frac{1}{\sigma^{np+p^2+p+2}} \cdot \frac{1}{|\Sigma|^{\frac{n+1}{2}}}$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i) \; -\frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t\Sigma^{-1}(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})\right\} \, d\Sigma d\sigma^2$$

$$\propto \left|\sum_{i=1}^{n}(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t\right|^{-\frac{n-p}{2}} \cdot \left[\sum_{i=1}^{n}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)\right]^{-\frac{np+p^2+p}{2}}. \tag{4.5}$$

Note that in the above derivation, we need $n \geq 2p$ to integrate out Σ. In the next section, we will prove that the RHS of (4.5) is proper if the number of full-rank X_i's is greater than $2p$.

4.3 The Proper of the Posterior Density

Suppose there are n_2 of X_i's that are of rank less than p, and these X_i's are the last n_2 of X_i's, i.e.,

$$r(X_1) = r(X_2) = \cdots = r(X_{n_1}) = p,$$
$$r(X_{n_1+1}) < p, \cdots, r(X_n) < p,$$

where $n_1 = n - n_2$ and $r(X_i)$ is the rank of X_i. Denote $\pi(\underset{\sim}{\beta},\ \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1} \mid data)$ be the marginal posterior for $\underset{\sim}{\beta},\ \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1}$. Then the following proposition will give an upper bound of $\pi(\underset{\sim}{\beta},\ \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1} \mid data)$. Here, we assume $n_2 \geq 1$, otherwise, we don't need the following proposition.

Proposition 4.1 *If $n_1 \geq 2p$, then*

$$\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1} \mid data)$$
$$\leq c\left|\sum_{i=1}^{n_1}(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t\right|^{-\frac{n_1-p}{2}} \cdot \left[\sum_{i=1}^{n_1}(\underset{\sim}{y}_i - X_i\beta_i)^t(\underset{\sim}{y}_i - X_i\beta_i)\right]^{-\frac{np+p^2+p}{2}}. \tag{4.6}$$

Proof: See Appendix D. ■

Now, we integrate out $\underset{\sim}{\beta}$ from (4.6).

Proposition 4.2 *If $n_1 \geq 2p+1$, then an upper bound of the marginal posterior density for $\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1}$ is given by*

$$\pi(\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1} \mid data) \leq c \cdot \left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t \right|^{-\frac{n_1-p-1}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i) \right]^{-\frac{np+p^2+p}{2}}, \tag{4.7}$$

where $\bar{\underset{\sim}{\beta}} = (1/n_1)\sum_{i=1}^{n_1} \underset{\sim}{\beta}_i$.

The proof is the same as for Proposition 4.1.

In order to prove the posterior proper for $\underset{\sim}{\beta}$, Σ, and σ^2, we need the following condition:

$$\text{There exist at least one } i \leq n_1 \text{ such that} \quad \underset{\sim}{y}_i \neq X_i(X_i^t X_i)^{-1} X_i^t \underset{\sim}{y}_i. \tag{4.8}$$

Under the above condition, it can be proved that there exist positive constant δ^* and δ, so that

$$\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i\beta_i)^t (\underset{\sim}{y}_i - X_i\beta_i) \geq \delta^* + \delta \sum_{i=1}^{n_1} \underset{\sim}{\beta}_i^t \underset{\sim}{\beta}_i. \tag{4.9}$$

By the above inequality and Proposition 4.2, it follows that

$$\pi(\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1} \mid data) \leq c \cdot \left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t \right|^{-\frac{n_1-p-1}{2}} \cdot \left[\delta^* + \delta \sum_{i=1}^{n_1} \underset{\sim}{\beta}_i^t \underset{\sim}{\beta}_i \right]^{-\frac{np+p^2+p}{2}}. \tag{4.10}$$

Note that

$$\begin{aligned} \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t &= (\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1})(I_{n_1} - \frac{1}{n_1} \underset{\sim}{1}\, \underset{\sim}{1}^t) \begin{pmatrix} \underset{\sim}{\beta}_1^t \\ \vdots \\ \underset{\sim}{\beta}_{n_1}^t \end{pmatrix}, \\ &= (\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1}) Q \begin{pmatrix} I_{n_1-1} & 0 \\ 0 & 0 \end{pmatrix} Q^t \begin{pmatrix} \underset{\sim}{\beta}_1^t \\ \vdots \\ \underset{\sim}{\beta}_{n_1}^t \end{pmatrix}, \end{aligned} \tag{4.11}$$

where I_{n_1} and I_{n_1-1} are the $n_1 \times n_1$ and $(n_1-1) \times (n_1-1)$ identity matrices, respectively, $\underset{\sim}{1}$ is an $n_1 \times 1$ vector with all elements equal 1, and Q is an $n_1 \times n_1$ orthogonal matrix. Consider the following transformation:

$$\eta \overset{\text{def}}{=} (\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1-1}, \underset{\sim}{\eta}_{n_1}) = (\underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_{n_1})Q, \tag{4.12}$$

where η is a $p \times n_1$ matrix. Thus, from (4.11) and (4.12),

$$\sum_{i=1}^{n_1}(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t = \sum_{i=1}^{n_1-1} \underset{\sim}{\eta}_i \underset{\sim}{\eta}_i^t, \tag{4.13}$$

$$\sum_{i=1}^{n_1} \underset{\sim}{\beta}_i^t \underset{\sim}{\beta}_i = \sum_{i=1}^{n_1} \underset{\sim}{\eta}_i^t \underset{\sim}{\eta}_i. \tag{4.14}$$

By Transformation (4.12) and Equations (4.10), (4.13) and (4.14), the bound of the marginal posterior density for $\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1}$ is

$$\pi(\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1} \mid data) \leq c \cdot \left| \sum_{i=1}^{n_1-1} \underset{\sim}{\eta}_i \underset{\sim}{\eta}_i^t \right|^{-\frac{n_1-p-1}{2}} \cdot \left[\delta^* + \delta \sum_{i=1}^{n_1} \underset{\sim}{\eta}_i^t \underset{\sim}{\eta}_i \right]^{-\frac{np+p^2+p}{2}}. \tag{4.15}$$

Integrating out $\underset{\sim}{\eta}_{n_1}$ from (4.15) yields

$$\pi(\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1-1} \mid data) \leq c \cdot \left| \sum_{i=1}^{n_1-1} \underset{\sim}{\eta}_i \underset{\sim}{\eta}_i^t \right|^{-\frac{n_1-p-1}{2}} \cdot \left[\delta^* + \delta \sum_{i=1}^{n_1-1} \underset{\sim}{\eta}_i^t \underset{\sim}{\eta}_i \right]^{-\frac{np+p^2}{2}}. \tag{4.16}$$

Therefore, the only thing left is to verify that $\pi(\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1-1} \mid data)$ is proper.

Denote $\eta_1 = (\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1-1})$, which is a $p \times (n_1 - 1)$ matrix. Then η_1 can be decomposed as follows:

$$\eta_1 = P \begin{pmatrix} \lambda_1 & & & 0 \\ & \ddots & & 0 \\ & & \lambda_p & 0 \end{pmatrix} \Gamma^t \overset{\text{def}}{=} PT\Gamma^t, \tag{4.17}$$

where P and Γ are $p \times p$ and $(n_1 - 1) \times (n_1 - 1)$ orthogonal matrices and $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. Rewrite Γ as $\Gamma = (\Gamma_1 \;\; \Gamma_2)$, where Γ_1 and Γ_2 are $(n_1 - 1) \times p$ and $(n_1 - 1) \times (n_1 - p - 1)$ matrices, respectively. Note that η_1 does not depend on Γ_2 since $\eta_1 = P \cdot \text{diag}(\lambda_1, \cdots, \lambda_p) \cdot \Gamma_1^t$.

In the following context, we assume that for every given Γ_1, let Γ_2 be a function of Γ_1 such that Γ is an orthogonal matrix.

Lemma 4.1 *The Jacobian of the orthogonal decomposition given by (4.17) is*

$$\left|\frac{\partial(\underset{\sim}{\eta}_1)}{\partial(P,T,\Gamma_1)}\right| = J(P,T,\Gamma_1) \propto \left(\prod_{i=1}^{p} \lambda_i\right)^{n_1-p-1} \prod_{1\le i<j\le p} |\lambda_i^2 - \lambda_j^2|. \tag{4.18}$$

Proof: See Appendix E. ■

Now we can complete the proof of the proper of the Jeffreys posterior.

Theorem 4.1 *If $n_1 \ge 2p+1$ and condition (4.8) is satisfied, then the Jeffreys posterior for the RCR model (1.1) is proper.*

Proof: Using the decomposition given in (4.17), Lemma 4.1 and Equation (4.16),

$$\begin{aligned}
&\int \pi(\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_{n_1-1} \mid data) d\underset{\sim}{\eta}_1 \cdots d\underset{\sim}{\eta}_{n_1-1} \\
\le\ & c \cdot \int \left(\lambda_1^2 \cdots \lambda_p^2\right)^{-\frac{n_1-p-1}{2}} \cdot \left[\delta^* + \delta \sum_{i=1}^{p} \lambda_i^2\right]^{-\frac{np+p^2}{2}} \cdot J(P,T,\Gamma_1) (\prod_{i=1}^{p} d\lambda_i)(P^t dP)(\Gamma_1^t d\Gamma_1) \\
\propto\ & \int \frac{\prod_{1\le i<j\le p} |\lambda_i^2 - \lambda_j^2|}{[\delta^* + \delta \sum_{i=1}^{p} \lambda_i^2]^{\frac{np+p^2}{2}}} (\prod_{i=1}^{p} d\lambda_i)(P^t dP)(\Gamma_1^t d\Gamma_1) < \infty
\end{aligned} \tag{4.19}$$

because $(P^t dP)$ and $(\Gamma_1^t d\Gamma_1)$ are finite **Haar** measures (for details, see, e.g., Muirhead [13], Chapter 2). ■

4.4 The posterior means and variances

The most interesting quantities in Bayesian inference are the posterior means and variances for the parameters of interest. In our RCR model (1.1), it is important to derive the posterior means and variances for $\underset{\sim}{\beta}$, Σ, and σ^2.

The following theorem is established for the Jeffreys posterior.

Theorem 4.2 *Under condition (4.8), the Jeffreys posterior means and variances exist for: (i) $\underset{\sim}{\beta}$, if $n_1 \ge 2p+3$; (ii) σ^2, if $n_1 \ge 2p+1$. Further, if $n_1 \ge 2p+2$, then*

$$E(\sigma_{ij}^{\frac{p+2}{2}} \mid data) < \infty. \tag{4.20}$$

Proof: See Appendix F. ■

Note that if $p \geq 2$, then from (4.20), the posterior means and variances of σ_{ij} exist when $n_1 \geq 2p + 2$.

Similarly, for the reference posterior, we have

Theorem 4.3 *Under condition (4.8) and $n_1 \geq 2p + 1$, the reference posteriors are proper. The reference posterior means and variances exist for: (i) $\underset{\sim}{\beta}$, if $n_1 \geq 2p + 3$; (ii) σ^2, if $n_1 \geq 2p + 1$. If $n_1 \geq 2p + 2$, then*

$$E(\sigma_{ij}^{\frac{p+1}{2}} \mid data) < \infty. \tag{4.21}$$

Also we have the following theorem for the "uniform" prior.

Theorem 4.4 *Under condition (4.8) and $n_1 \geq 2p + 2$, the posterior corresponds to the "uniform" prior is proper. The posterior means and variances exist for: (i) $\underset{\sim}{\beta}$, if $n_1 \geq 2p + 4$; (ii) σ^2, if $n_1 \geq 2p + 2$.*

Note that for the "uniform" prior, the conditions in Theorem 4.4 are also necessary conditions for the posterior to be proper and the posterior means to be existent if $n_2 = 0$.

5 Computation of Posterior Expectations

5.1 Transformation

In computing the posterior expectations in Section 4.4, it will be convenient to use transformations: (i) $\underset{\sim}{\beta} = \underset{\sim}{\beta}$, $\Lambda = \Sigma/\sigma^2$ and $\sigma^2 = \sigma^2$; then (ii) $\Lambda^* = \log \Lambda$ or $\Lambda = e^{\Lambda^*}$. The advantages of the above transformations are that the conditional distribution for σ^2 given $(\underset{\sim}{\beta},\ \Lambda)$ is an inverse Gamma distribution (see (5.2) below) and that Λ^* is an unconstrained symmetric matrix while Λ is a positive definite matrix. It has been shown in Yang and Berger [17] that the Jacobian of transformation (ii) is

$$\left|\frac{\partial \Lambda}{\partial \Lambda^*}\right| \propto \frac{|\Lambda| \prod_{i<j}(d_i - d_j)}{\prod_{i<j}(d_i^* - d_j^*)}, \tag{5.1}$$

where d_i and d_i^* are eigenvalues of Λ and Λ^*, respectively.

Using prior $\pi(\underset{\sim}{\beta}, \Lambda, \sigma^2)$, combining with the likelihood function (2.2) and the Jacobian (5.1), the posterior of $(\underset{\sim}{\beta}, \Lambda^*, \sigma^2)$ thus is

$$\pi(\underset{\sim}{\beta}, \Lambda^*, \sigma^2 \mid data) \propto \pi(\underset{\sim}{\beta}, e^{\Lambda^*}, \sigma^2) \cdot \sigma^{-\sum_{i=1}^n t_i} \cdot \prod_{i=1}^n \left| X_i e^{\Lambda^*} X_i^t + I_i \right|^{-\frac{1}{2}}$$
$$\cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta})^t (X_i e^{\Lambda^*} X_i^t + I_i)^{-1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}) \right\} \cdot \frac{|e^{\Lambda^*}| \prod_{i<j}(d_i - d_j)}{\prod_{i<j}(d_i^* - d_j^*)}. \quad (5.2)$$

5.2 Hybrid Markov Chain Sampling

A hybrid Markov Chain Sampling scheme will be used for computing the posterior expectations in Section 4.4 since these posterior expectations are not available in closed form. The hybrid scheme, which generates a dependent sample $\{(\underset{\sim}{\beta}_i, \Lambda_i^*, \sigma_i^2),\ i \geq 0\}$ from the posterior (5.2), has three major components: grouping, an approximate Gibbs step, and Hit-and-Run generation. The parameters are grouped as (i) $\underset{\sim}{\beta}$, (ii) Λ^* and (iii) σ^2. Then, the approximate Gibbs step proceeds as follows:

Step 0. Choose a starting point $(\underset{\sim}{\beta}_0,\ \Lambda_0^*,\ \sigma_0^2)$, and set $k = 0$.

Step 1. Generate $\underset{\sim}{\beta}_{(k+1)}$ from the conditional posterior distribution

$$\pi(\underset{\sim}{\beta} \mid \Lambda_k^*,\ \sigma_k^2,\ data),$$

which is a multivariate Normal distribution.

Step 2. Generate $\Lambda_{(k+1)}^*$ from the (approximately) conditional posterior distribution

$$\pi(\Lambda^* \mid \underset{\sim}{\beta}_{(k+1)},\ \sigma_k^2,\ data).$$

Step 3. Generate $\sigma_{(k+1)}^2$ from the conditional posterior distribution

$$\pi(\sigma^2 \mid \underset{\sim}{\beta}_{(k+1)},\ \Lambda_k^*,\ data),$$

which is an inverse Gamma distribution.

Step 4. Set $k = k + 1$ and go to Step 1.

For Step 2, we approximately sample from the conditional posterior distribution using one-step of a Metropolisized Hit-and-Run sampler. The procedure proceeds as follows:

(i) Generate a random direction (symmetric) matrix T, defined by $T = Z/\sqrt{\sum_{i \leq j} z_{ij}^2}$, where $z_{ij} \overset{i.i.d}{\sim} N(0,\ 1)$, $i \leq j$, and Z is the symmetric matrix with the $(i,\ j)^{\text{th}}$ element z_{ij}, $i \leq j$.

(ii) Generate $\lambda \sim N(0,\ 1)$.

(iii) Set $Y = \Lambda_k^* + \lambda T$. Then set

$$\Lambda_{k+1}^* = \begin{cases} Y, & \text{with probability } \min(1, \pi(Y|\underset{\sim}{\beta}_{(k+1)}, \sigma_k^2, data)/\pi(\Lambda_k^*|\underset{\sim}{\beta}_{(k+1)}, \sigma_k^2, data)) \\ \Lambda_k^*, & \text{otherwise.} \end{cases} \tag{5.3}$$

After a sufficiently large sample $(\underset{\sim}{\beta}_1,\ \Lambda_1^*,\ \sigma_1^2)$, $(\underset{\sim}{\beta}_2,\ \Lambda_2^*,\ \sigma_2^2)$, $\ldots$, $(\underset{\sim}{\beta}_m,\ \Lambda_m^*,\ \sigma_m^2)$ has been generated, the posterior expectation is approximated by

$$E^{\pi(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2|data)} h(\underset{\sim}{\beta},\ \Sigma,\ \sigma^2) \approx \frac{1}{m}\sum_{k=1}^{m} h(\underset{\sim}{\beta}_k,\ \exp(\Lambda_k^*)\cdot\sigma_k^2,\ \sigma_k^2),$$

where h is the function of interest. General discussion about this hybrid algorithm can be found in Berger and Chen [5].

The above hybrid Markov chain scheme was programmed in double precision Fortran-77 using the IMSL subroutines. It is available from the authors upon request.

5.3 Example

Consider the Model:

$$y_{ij} = \alpha_i + x_{ij}\beta_i + \epsilon_{ij}, \quad j = 1, \ldots, t_i, \quad i = 1, \ldots, n. \tag{5.4}$$

where

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right) \text{ and independently } \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0,\ 1).$$

The simulated data is listed in Table I. Note that the first two design matrices are singular.

Using the reference prior w.r.t. the group ordering $\{\underset{\sim}{\beta}, \sigma^2, \Lambda\}$, i.e., $|G\sum_{i=1}^n (B_i \otimes B_i)G^t|^{\frac{1}{2}}/\sigma^2$, the posterior means are

$$\hat{\underset{\sim}{\beta}} = \begin{pmatrix} 1.864 \\ 4.036 \end{pmatrix}, \quad \hat{\sigma}^2 = 1.199, \quad \hat{\Sigma} = \begin{pmatrix} 1.631 & 1.020 \\ 1.020 & 2.958 \end{pmatrix},$$

which are not off the true parameters very much as the sample size $n = 10$.

Table I.

Simulated Data from Model (5.4)

	x_{ij}					
subject	0	0	1	1	2	2
1	-0.497					
2	3.976	3.043				
3	-0.205	1.004	3.922			
4	3.409	2.279	7.128	7.307		
5	0.462	2.971	5.143	7.248	10.836	
6	2.608	1.988	6.144	5.058	11.272	10.126
7	2.343	3.298	7.724	4.367	10.595	9.557
8	2.234	2.822	3.787	4.185	8.379	8.993
9	0.172	0.627	4.505	4.741	7.110	4.928
10	2.440	2.726	7.026	10.972	14.947	15.790

6 Comments and Generalizations

6.1 Comments

For the RCR model (1.1), it is difficult to do comparisons between noninformative priors, partly because there is no unique way to compare noninformative priors, and there are also too many choices in the parameter setting and design matrix setting. However, in terms of MSE criteria for the corresponding posterior mean estimators, we empirically observed that (i) both reference priors (3.3) and Jeffreys prior (3.1) result in similar Bayes estimators for $\underset{\sim}{\beta}$ and σ^2, and both produce smaller MSEs than "uniform" prior for Bayes estimators of $\underset{\sim}{\beta}$ and σ^2; (ii) Jeffreys prior (3.1) produces Bayes estimator for Σ with the smallest MSE, and "Uniform" prior produces Bayes estimator for Σ with the largest MSE.

For estimating variance or covariance matrix, it may not be a good idea to use square loss. It may be more desirable to use some invariant losses, e.g.,

$$L(\hat{\sigma}^2,\ \sigma^2) = (\hat{\sigma}^2 - \sigma^2)^2/\sigma^4. \tag{6.1}$$

Under such a loss, reference priors (3.3) would produce Bayes estimator with the smallest risk.

In conjunction with the above discussion and the consideration of simplicity of a prior, the reference prior w.r.t. the group ordering $\{\underset{\sim}{\beta}, \sigma^2, \Lambda\}$, i.e., $|G\sum_{i=1}^{n}(B_i \otimes B_i)G^t|^{\frac{1}{2}}/\sigma^2$, would be preferred. Further serious simulation study to this prior is also recommended.

6.2 Generalization of RCR Model (1.1)

The RCR model (1.1) can be generalized to the following mixed linear model, which was previously studied by Swamy [14] (pp.143),

$$\underset{\sim}{y}_i = Z_i\underset{\sim}{\alpha} + X_i\underset{\sim}{\beta}_i + \varepsilon_i, \tag{6.2}$$

where $\underset{\sim}{y}_i$ is a $t_i \times 1$ vector of observations. X_i and Z_i are $t_i \times p$ and $t_i \times q$ design matrix, $\underset{\sim}{\alpha}$ is a $q \times 1$ vector of fixed effects. $\underset{\sim}{\beta}_i$ is a $p \times 1$ vector of random coefficients for the ith experimental subject and ε_i is a vector of errors for $i = 1, 2, \cdots, n$. $\underset{\sim}{\beta}_i$ and ε_i are independently distributed as $MVN(\underset{\sim}{\beta},\ \Sigma)$ and $MVN(\underset{\sim}{0},\ \sigma^2 I_i)$, respectively.

The computation of the Fisher information matrix for the model (6.2) is similarly to that in Proposition 2.1. The only difference to (2.12) is that $I_i(\underset{\sim}{\alpha},\ \underset{\sim}{\beta}) = (Z_i\ X_i)^t(X_i\Sigma X_i^t + \sigma^2 I_i)^{-1}(Z_i\ X_i)$. The Jeffreys prior for model (6.2) is thus similar to that in (3.1), and reference prior for model (6.2) is the same as (3.3).

Further, Theorems 4.1, 4.2, 4.3 and 4.4 also hold for model (6.2), provided that the condition (4.8) is changed to that

$$\text{There exist at least one } i \leq n_1 \text{ such that} \quad \underset{\sim}{y}_i \neq (Z_i\ X_i)[(Z_i\ X_i)^t(Z_i\ X_i)]^{-1}(Z_i\ X_i)^t\underset{\sim}{y}_i. \tag{6.3}$$

Finally, the computation of the posterior expectation can also be carried out accordingly.

Appendix A: Proof of Proposition 2.1

Lemma 2.1 yields

$$\begin{aligned} I_i(\underset{\sim}{\beta}) &= \left[\frac{\partial(X_i\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\right] (X_i\Sigma X_i^t + \sigma^2 I_i)^{-1} \left[\frac{\partial(X_i\underset{\sim}{\beta})}{\partial\underset{\sim}{\beta}}\right]^t \\ &= X_i^t(X_i\Sigma X_i^t + \sigma^2 I_i)^{-1} X_i \end{aligned} \tag{A.1}$$

and

$$I_i(\Sigma,\sigma^2) = \left[\frac{\partial vec(X_i\Sigma X_i^t + \sigma^2 I_i)}{\partial vec(\Sigma,\sigma^2)}\right] \cdot \left[\Sigma_i^{*-1} \otimes \Sigma_i^{*-1}\right] \cdot \left[\frac{\partial vec(X_i\Sigma X_i^t + \sigma^2 I_i)}{\partial vec(\Sigma,\sigma^2)}\right]^t, \tag{A.2}$$

where $\Sigma_i^* = X_i\Sigma X_i^t + \sigma^2 I_i$. Denote $\Sigma = (\sigma_{lm})_{p\times p}$. Then

$$\frac{\partial vec(X_i\Sigma X_i^t + \sigma^2 I_i)}{\partial \sigma_{lm}} = vec\left(X_i \frac{\partial\Sigma}{\partial\sigma_{lm}} X_i^t\right) = vec(X_i\Delta_{lm}X_i^t), \tag{A.3}$$

where $\Delta_{lm} = e_l e_m^t + e_m e_l^t$ if $l < m$ and $\Delta_{mm} = e_m e_m^t$. Here $e_l = (0,\cdots,0,1,0,\cdots,0)^t$ denotes a column vector with 1 in the l^{th} row and 0 elsewhere. Then (A.3) and Result 2.1 imply

$$\begin{aligned} I_i(\sigma_{lm},\sigma_{hk}) &= \left[vec(X_i\Delta_{lm}X_i^t)\right]^t \cdot \left(\Sigma_i^{*-1} \otimes \Sigma_i^{*-1}\right) \cdot \left[vec(X_i\Delta_{hk}X_i^t)\right] \\ &= tr\left[\Sigma_i^{*-1} \cdot X_i\Delta_{hk}X_i^t \cdot \Sigma_i^{*-1} \cdot X_i\Delta_{lm}X_i^t\right] \\ &= tr\left[A_i \cdot \Delta_{hk} \cdot A_i \cdot \Delta_{lm}\right] = (vec(\Delta_{lm}))^t \, (A_i \otimes A_i) \, (vec(\Delta_{hk})). \end{aligned} \tag{A.4}$$

Therefore (A.4) results in

$$I_i(\Sigma) = G(A_i \otimes A_i)G^t. \tag{A.5}$$

Since

$$\frac{\partial vec(X_i\Sigma X_i^t + \sigma^2 I_i)}{\partial \sigma^2} = vec(I_i),$$

then using Result 2.1,

$$\begin{aligned} I_i(\sigma_{lm},\sigma^2) &= \left[vec(X_i\Delta_{lm}X_i^t)\right]^t \cdot \left(\Sigma_i^{*-1} \otimes \Sigma_i^{*-1}\right) \cdot [vec(I_i)] \\ &= tr\left[\Sigma_i^{*-1} \cdot I_i \cdot \Sigma_i^{*-1} \cdot X_i\Delta_{lm}X_i^t\right] = [vec(\Delta_{lm})]^t \cdot (B_i \otimes B_i) \cdot [vec(I_i)], \end{aligned} \tag{A.6}$$

where $B_i = (X_i \Sigma X_i^t + \sigma^2 I_i)^{-1} X_i$. Thus, Result 2.2 and (A.6) result in

$$I_i(\Sigma, \sigma^2) = \left[\frac{\partial vec(\Sigma)}{\partial vecp(\Sigma)}\right] \cdot (B_i \otimes B_i) \cdot [vec(I_i)] = Gvec(C_i). \tag{A.7}$$

Using Result 2.1 again,

$$I_i(\sigma^2) = [vec(I_i)]^t \cdot \left(\Sigma_i^{*-1} \otimes \Sigma_i^{*-1}\right) \cdot [vec(I_i)] = tr\left((X_i \Sigma X_i^t + \sigma^2 I_i)^{-2}\right). \tag{A.8}$$

Therefore Proposition 2.1 follows from (A.1), (A.5), (A.7), and (A.8). ∎

Appendix B: Proof of Theorem 3.1

The detail of the algorithm to compute the grouped ordering reference prior can be found in Berger and Bernardo [4]. We will compute the reference prior only for the group ordering ($\underset{\sim}{\beta}$, σ^2, Λ) here. The reference priors for other group orderings can be computed similarly.

Following the notations in Berger and Bernardo [4], the functions h_j, which are needed to calculate the reference prior, are

$$\begin{aligned}
|h_3| &= |G(\sum_{i=1}^n B_i \otimes B_i) G^t/2|, \\
|h_2| &= \left|\sum_{i=1}^n t_i - [Gvec(\sum_{i=1}^n B_i)]^t [G(\sum_{i=1}^n B_i \otimes B_i) G^t]^{-1} [Gvec(\sum_{i=1}^n B_i)]\right| / (2\sigma^4), \\
|h_1| &= |\sum_{i=1}^n B_i| / \sigma^{2p},
\end{aligned}$$

Then, the conditional prior of Λ given $\underset{\sim}{\beta}$ and σ^2 is

$$\pi_l(\Lambda | \underset{\sim}{\beta}, \sigma^2) \propto |h_3|^{\frac{1}{2}} = |G(\sum_{i=1}^n B_i \otimes B_i) G^t|^{\frac{1}{2}}.$$

Hence $E[\log |h_2| | \underset{\sim}{\beta}, \sigma^2] = -\log \sigma^4 + c$. So the conditional prior of Λ and σ^2 given $\underset{\sim}{\beta}$ is

$$\pi_l(\sigma^2, \Lambda | \underset{\sim}{\beta}) \propto |G(\sum_{i=1}^n B_i \otimes B_i) G^t|^{\frac{1}{2}} / \sigma^2.$$

It implies that $E[\log|h_1||(\underset{\sim}{\beta})] = c$ and

$$\pi_l(\underset{\sim}{\beta}, \sigma^2, \Lambda) \propto |G(\sum_{i=1}^{n} B_i \otimes B_i)G^t|^{\frac{1}{2}}/\sigma^2.$$

The conclusion of the Theorem thus follows. ■

Appendix C: Proof of Proposition 3.1

(3.4) is trivial. For (3.5), since $\Sigma \geq 0$, then Σ can be decomposed as

$$\Sigma = Q \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^* \end{pmatrix} Q^t,$$

where λ is the largest eigenvalue of Σ and Q is a $p \times p$ orthogonal matrix. By denote $X_i Q = \left(\underset{\sim}{h}_1^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right)$, where $\underset{\sim}{h}_j^{(i)}$ is a $t_i \times 1$ vector, it follows that

$$X_i \Sigma X_i^t + \sigma^2 I_i \geq \left(\underset{\sim}{h}_1^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right) \begin{pmatrix} \lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \begin{pmatrix} \underset{\sim}{h}_1^{(i)t} \\ \vdots \\ \underset{\sim}{h}_p^{(i)t} \end{pmatrix} + \sigma^2 I_i = \lambda \underset{\sim}{h}_1^{(i)} \underset{\sim}{h}_1^{(i)t} + \sigma^2 I_i.$$

Therefore,

$$Q^t X_i^t \left(X_i \Sigma X_i^t + \sigma^2 I_i\right)^{-1} X_i Q \leq \begin{pmatrix} \underset{\sim}{h}_1^{(i)t} \\ \vdots \\ \underset{\sim}{h}_p^{(i)t} \end{pmatrix} \cdot \frac{1}{\sigma^2} \cdot \left[I_i - \frac{\underset{\sim}{h}_1^{(i)} \underset{\sim}{h}_1^{(i)t}}{\underset{\sim}{h}_1^{(i)t} \underset{\sim}{h}_1^{(i)} + \sigma^2/\lambda} \right] \cdot \left(\underset{\sim}{h}_1^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right). \quad \text{(C.1)}$$

Denote the right hand side of (C.1) as $\begin{pmatrix} h_{11}^{(i)} & * \\ * & H_{p-1,p-1}^{(i)} \end{pmatrix}$. Then

$$h_{11}^{(i)} = \underset{\sim}{h}_1^{(i)t} \cdot \frac{1}{\sigma^2} \cdot \left[I_i - \frac{\underset{\sim}{h}_1^{(i)} \underset{\sim}{h}_1^{(i)t}}{\underset{\sim}{h}_1^{(i)t} \underset{\sim}{h}_1^{(i)} + \sigma^2/\lambda} \right] \cdot \underset{\sim}{h}_1^{(i)} = \frac{1}{\lambda} \cdot \frac{\underset{\sim}{h}_1^{(i)t} \underset{\sim}{h}_1^{(i)}}{\underset{\sim}{h}_1^{(i)t} \underset{\sim}{h}_1^{(i)} + \sigma^2/\lambda} \leq \frac{1}{\lambda}; \quad \text{(C.2)}$$

and

$$H^{(i)}_{p-1,p-1} \quad \leq \quad \begin{pmatrix} \underset{\sim}{h}_2^{(i)^t} \\ \vdots \\ \underset{\sim}{h}_p^{(i)^t} \end{pmatrix} \cdot \frac{I_i}{\sigma^2} \cdot \left(\underset{\sim}{h}_2^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right). \tag{C.3}$$

By Result 2.3, (C.1), (C.2) and (C.3),

$$\begin{aligned} \left|\sum_{i=1}^n A_i\right| \quad &\leq \quad \left|\sum_{i=1}^n h_{11}^{(i)}\right| \cdot \left|\sum_{i=1}^n H^{(i)}_{p-1,p-1}\right| \leq \frac{n}{\lambda} \cdot \left|\sum_{i=1}^n \begin{pmatrix} \underset{\sim}{h}_2^{(i)^t} \\ \vdots \\ \underset{\sim}{h}_p^{(i)^t} \end{pmatrix} \cdot \frac{I_i}{\sigma^2} \cdot \left(\underset{\sim}{h}_2^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right)\right| \\ &= \quad \frac{n}{\lambda\sigma^{2(p-1)}} \cdot \left|\sum_{i=1}^n \begin{pmatrix} \underset{\sim}{h}_2^{(i)^t} \\ \vdots \\ \underset{\sim}{h}_p^{(i)^t} \end{pmatrix} \left(\underset{\sim}{h}_2^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)}\right)\right| \leq \frac{c}{\lambda\sigma^{2(p-1)}}, \end{aligned} \tag{C.4}$$

Note that the last step of (C.4) follows from the fact that the elements of $X_iQ = (\underset{\sim}{h}_1^{(i)}, \cdots, \underset{\sim}{h}_p^{(i)})$ are uniformly bounded. ■

Appendix D: Proof of Proposition 4.1

From (4.5),

$$\pi(\underset{\sim}{\beta}, \underset{\sim}{\beta}_1, \cdots, \underset{\sim}{\beta}_n \mid data) \leq c \cdot \left|\sum_{i=1}^n (\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t\right|^{-\frac{n-p}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)\right]^{-\frac{np+p^2+p}{2}}. \tag{D.1}$$

Now, we first integrate out $\underset{\sim}{\beta}_n$ from the RHS of (D.1). By writing $A = \sum_{i=1}^{n-1}(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t$ and $\underset{\sim}{\beta}_n^* = A^{-\frac{1}{2}}(\underset{\sim}{\beta}_n - \underset{\sim}{\beta})$, we have

$$\begin{aligned} &\int \left|\sum_{i=1}^n (\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t\right|^{-\frac{n-p}{2}} d\underset{\sim}{\beta}_n \\ = \quad &|A|^{-\frac{n-p}{2}} \int \left|A^{-\frac{1}{2}}(\underset{\sim}{\beta}_n - \underset{\sim}{\beta})(\underset{\sim}{\beta}_n - \underset{\sim}{\beta})^t A^{-\frac{1}{2}} + I\right|^{-\frac{n-p}{2}} d\underset{\sim}{\beta}_n \\ = \quad &|A|^{-\frac{n-p-1}{2}} \int \left|\underset{\sim}{\beta}_n^* \underset{\sim}{\beta}_n^{*t} + I\right|^{-\frac{n-p}{2}} d\underset{\sim}{\beta}_n^* \end{aligned}$$

$$
\begin{aligned}
&= |A|^{-\frac{n-p-1}{2}} \int (1+\underset{\sim}{\beta}_n^{*t}\underset{\sim}{\beta}_n^{*})^{-\frac{n-p}{2}} d\underset{\sim}{\beta}_n^{*} \\
&\propto \left|\sum_{i=1}^{n-1}(\underset{\sim}{\beta}_i-\underset{\sim}{\beta})(\underset{\sim}{\beta}_i-\underset{\sim}{\beta})^t\right|^{-\frac{n-p-1}{2}},
\end{aligned} \tag{D.2}
$$

Repeating the above procedure to integrate out $\underset{\sim}{\beta}_{n_1+1},\cdots,\underset{\sim}{\beta}_{n-1}$ gives Proposition 4.1. Note that $\int(1+\underset{\sim}{\beta}^{*t}\underset{\sim}{\beta}^{*})^{-\frac{n-k}{2}}d\underset{\sim}{\beta}^{*}<\infty$ if $n-k>p$. ∎

Appendix E: Proof of Lemma 4.1

Using the exterior product method in Muirhead [13] (Chapter 2) and $\eta_1 = PT\Gamma^t$,

$$d\eta_1 = dP\cdot T\Gamma^t + P\cdot dT\cdot\Gamma^t + PT\cdot d\Gamma^t.$$

Thus, by the fact $d\Gamma^t\Gamma = -\Gamma^t d\Gamma$,

$$P^t d\eta_1\Gamma = P^t dP\cdot T - T\cdot\Gamma^t d\Gamma + dT. \tag{E.1}$$

According to Chapter 2 of Muirhead [13], the exterior product of elements in the matrix on the left hand side of (E.1) is

$$\left|P^t\right|^p\cdot|\Gamma|^{n_1-1}\cdot(d\eta_1) = (d\eta_1). \tag{E.2}$$

Write $P=(\underset{\sim}{P}_1,\underset{\sim}{P}_2,\cdots,\underset{\sim}{P}_p)$ and $\Gamma=(\gamma_1,\gamma_2,\cdots,\gamma_{n_1-1})$. Then

$$
\begin{aligned}
P^t dPT &= \begin{pmatrix} \underset{\sim}{P}_1^t \\ \underset{\sim}{P}_2^t \\ \vdots \\ \underset{\sim}{P}_p^t \end{pmatrix} (d\underset{\sim}{P}_1,\cdots,d\underset{\sim}{P}_p) \begin{pmatrix} \lambda_1 & & & 0 \\ & \ddots & & 0 \\ & & \lambda_p & 0 \end{pmatrix} \\
&= \begin{pmatrix}
0 & -\lambda_2\underset{\sim}{P}_2^t d\underset{\sim}{P}_1 & \cdots & -\lambda_p\underset{\sim}{P}_p^t d\underset{\sim}{P}_1 & 0 \\
\lambda_1\underset{\sim}{P}_2^t d\underset{\sim}{P}_1 & 0 & \cdots & -\lambda_p\underset{\sim}{P}_p^t d\underset{\sim}{P}_2 & 0 \\
\lambda_1\underset{\sim}{P}_3^t d\underset{\sim}{P}_1 & \lambda_2\underset{\sim}{P}_3^t d\underset{\sim}{P}_2 & \cdots & -\lambda_p\underset{\sim}{P}_p^t d\underset{\sim}{P}_3 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\lambda_1\underset{\sim}{P}_p^t d\underset{\sim}{P}_1 & \lambda_2\underset{\sim}{P}_p^t d\underset{\sim}{P}_2 & \cdots & 0 & 0
\end{pmatrix}.
\end{aligned} \tag{E.3}
$$

Similarly, we have

$$T\Gamma^t d\Gamma = \begin{pmatrix} 0 & -\lambda_1 \gamma_2^t d\gamma_1 & \cdots & -\lambda_1 \gamma_p^t d\gamma_1 & -\lambda_1 \gamma_{p+1}^t d\gamma_1 & \cdots & -\lambda_1 \gamma_{n_1-1}^t d\gamma_1 \\ \lambda_2 \gamma_2^t d\gamma_1 & 0 & \cdots & -\lambda_2 \gamma_p^t d\gamma_2 & -\lambda_2 \gamma_{p+1}^t d\gamma_2 & \cdots & -\lambda_2 \gamma_{n_1-1}^t d\gamma_2 \\ \lambda_3 \gamma_3^t d\gamma_1 & \lambda_3 \gamma_3^t d\gamma_2 & \cdots & -\lambda_3 \gamma_p^t d\gamma_3 & -\lambda_3 \gamma_{p+1}^t d\gamma_3 & \cdots & -\lambda_3 \gamma_{n_1-1}^t d\gamma_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \lambda_p \gamma_p^t d\gamma_1 & \lambda_p \gamma_p^t d\gamma_2 & \cdots & 0 & -\lambda_p \gamma_{p+1}^t d\gamma_p & \cdots & -\lambda_p \gamma_{n_1-1}^t d\gamma_p \end{pmatrix}; \quad \text{(E.4)}$$

and

$$dT = \begin{pmatrix} d\lambda_1 & & 0 \\ & \ddots & 0 \\ & d\lambda_p & 0 \end{pmatrix}. \quad \text{(E.5)}$$

From (E.3), (E.4) and (E.5), the exterior product of the right hand side of (E.1) is

$$\bigwedge_{i=1}^{p} d\lambda_i \bigwedge_{1\le i<j\le p} \left[(\lambda_i P_j^t dP_i - \lambda_j \gamma_j^t d\gamma_i) \bigwedge (-\lambda_j P_j^t dP_i + \lambda_i \gamma_j^t d\gamma_i)\right] \bigwedge_{\substack{1\le i\le p \\ p+1\le j\le n_1-1}} \lambda_i \gamma_j^t d\gamma_i$$

$$= \left(\prod_{i=1}^{p} \lambda_i\right)^{n_1-p-1} \prod_{1\le i<j\le p} |\lambda_i^2 - \lambda_j^2| \bigwedge_{i=1}^{p} d\lambda_i \bigwedge_{1\le i<j\le p} \left(P_j^t dP_i \bigwedge \gamma_j^t d\gamma_i\right) \bigwedge_{\substack{1\le i\le p \\ p+1\le j\le n_1-1}} \gamma_j^t d\gamma_i . \quad \text{(E.6)}$$

(E.2) and (E.6) yield

$$\left|\frac{\partial(\eta_1)}{\partial(P,T,\Gamma_1)}\right| \propto \left(\prod_{i=1}^{p} \lambda_i\right)^{n_1-p-1} \cdot \prod_{1\le i<j\le p} |\lambda_i^2 - \lambda_j^2|.$$

■

Appendix F: Proof of Theorem 4.2

It is straightforward to prove that if $n_1 \geq 2p+1$, then the posterior mean and variance of σ^2 exist.

In order to prove the posterior variances of β exist, it suffices to prove that

$$E(\beta^t \beta \mid data) < \infty. \quad \text{(F.1)}$$

Using Proposition 4.1, (F.1) is true if

$$\int \underset{\sim}{\beta}^t \underset{\sim}{\beta} \left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \underset{\sim}{\beta})(\underset{\sim}{\beta}_i - \underset{\sim}{\beta})^t \right|^{-\frac{n_1-p}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i) \right]^{-\frac{np+p^2+p}{2}} d\underset{\sim}{\beta} \prod_{i=1}^{p} d\underset{\sim}{\beta}_i < \infty. \quad \text{(F.2)}$$

Letting $A = \frac{1}{n_1} \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t$ and $\underset{\sim}{\beta}^* = A^{-\frac{1}{2}}(\underset{\sim}{\beta} - \bar{\underset{\sim}{\beta}})$, we have

$$\underset{\sim}{\beta}^t \underset{\sim}{\beta} = \left(\underset{\sim}{\beta}^{*t} A^{\frac{1}{2}} + \bar{\underset{\sim}{\beta}}^t \right) \left(A^{\frac{1}{2}} \underset{\sim}{\beta}^* + \bar{\underset{\sim}{\beta}} \right) = \underset{\sim}{\beta}^{*t} A \underset{\sim}{\beta}^* + 2 \bar{\underset{\sim}{\beta}}^t A^{\frac{1}{2}} \underset{\sim}{\beta}^* + \bar{\underset{\sim}{\beta}}^t \bar{\underset{\sim}{\beta}}.$$

Then, the Cauchy-Schwartz inequality yields

$$n_1 \underset{\sim}{\beta}^{*t} A \underset{\sim}{\beta}^* = \sum_{i=1}^{n_1} \left[\underset{\sim}{\beta}^{*t} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}}) \right]^2 \le \underset{\sim}{\beta}^{*t} \underset{\sim}{\beta}^* \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}}). \quad \text{(F.3)}$$

Denote the LHS of (F.2) to be I^*. Then by (F.3), using the same technique as in the proof of Proposition 4.1,

$$I^* \le I_1^* + I_2^* + I_3^*, \quad \text{(F.4)}$$

where

$$\begin{aligned} I_1^* &= \int \frac{(1/n_1) \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}}) \prod_{i=1}^{p} d\underset{\sim}{\beta}_i}{\left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t \right|^{\frac{n_1-p-1}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i) \right]^{\frac{np+p^2+p}{2}}} \\ &\quad \cdot \int \frac{\underset{\sim}{\beta}^{*t} \underset{\sim}{\beta}^* d\underset{\sim}{\beta}^*}{(1 + \underset{\sim}{\beta}^{*t} \underset{\sim}{\beta}^*)^{\frac{n_1-p}{2}}}, \quad \text{(F.5)} \\ I_2^* &= \int \left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t \right|^{-\frac{n_1-p-1}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i) \right]^{-\frac{np+p^2+p}{2}} \prod_{i=1}^{p} d\underset{\sim}{\beta}_i \\ &\quad \cdot \int \frac{2 \bar{\underset{\sim}{\beta}}^t A^{\frac{1}{2}} \underset{\sim}{\beta}^*}{(1 + \underset{\sim}{\beta}^{*t} \underset{\sim}{\beta}^*)^{\frac{n_1-p}{2}}} d\underset{\sim}{\beta}^* = 0, \quad \text{(F.6)} \\ I_3^* &= \int \frac{\bar{\underset{\sim}{\beta}}^t \bar{\underset{\sim}{\beta}} \prod_{i=1}^{p} d\underset{\sim}{\beta}_i}{\left| \sum_{i=1}^{n_1} (\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t \right|^{\frac{n_1-p-1}{2}} \cdot \left[\sum_{i=1}^{n_1} (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i)^t (\underset{\sim}{y}_i - X_i \underset{\sim}{\beta}_i) \right]^{\frac{np+p^2+p}{2}}} \\ &\quad \int \left(1 + \underset{\sim}{\beta}^{*t} \underset{\sim}{\beta}^* \right)^{-\frac{n_1-p}{2}} d\underset{\sim}{\beta}^*. \quad \text{(F.7)} \end{aligned}$$

Since

$$\frac{(1/n_1)\sum_{i=1}^{n_1}(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})^t(\underset{\sim}{\beta}_i - \bar{\underset{\sim}{\beta}})}{[\sum_{i=1}^{n_1}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)]^{\frac{np+p^2+p}{2}}} \le c \cdot \left[\sum_{i=1}^{n_1}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)\right]^{-\frac{np+p^2+p-4}{2}},$$

$$\frac{\bar{\underset{\sim}{\beta}}^t\bar{\underset{\sim}{\beta}}}{[\sum_{i=1}^{n_1}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)]^{\frac{np+p^2+p}{2}}} \le c \cdot \left[\sum_{i=1}^{n_1}(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)^t(\underset{\sim}{y}_i - X_i\underset{\sim}{\beta}_i)\right]^{-\frac{np+p^2+p-4}{2}},$$

and when $n_1 \ge 2p+3$,

$$\int \frac{\underset{\sim}{\beta}^{*t}\underset{\sim}{\beta}^* d\underset{\sim}{\beta}^*}{(1+\underset{\sim}{\beta}^{*t}\underset{\sim}{\beta}^*)^{\frac{n_1-p}{2}}} < \infty, \text{ and } \int \left(1+\underset{\sim}{\beta}^{*t}\underset{\sim}{\beta}^*\right)^{-\frac{n_1-p}{2}} d\underset{\sim}{\beta}^* < \infty,$$

then $I_1^* < \infty$, and $I_3^* < \infty$. Therefore $I^* < \infty$.

For proving (4.20), it suffices to prove that

$$E(\lambda^{\frac{p+2}{2}} \mid data) < \infty, \tag{F.8}$$

where λ is the largest eigenvalue of Σ. Using the upper bound of the Jeffreys prior, which is given by (3.7) in Proposition 3.2, (F.8) can be proven in the same way as we prove Theorem 4.1. ■

Acknowledgement

We are grateful to the authors' advisor Professor James O. Berger for many valuable comments and suggestions.

References

[1] Berger, J. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84** 200-207.

[2] Berger, J. and Bernardo, J.M. (1992a). Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, (P. Goel and N.S. Iyengar, Eds.), pp.177-194. Springer Verlag, New York.

[3] Berger, J. and Bernardo, J.M. (1992b). Ordered group reference priors with application to a multinomial problem. *Biometrika* **79** 25-37.

[4] Berger, J. and Bernardo, J.M. (1992c). On the development of the reference prior method. In *Bayesian Statistics 4*, (J.M.Bernardo, J.O.Berger, D.V.Lindley, and A.F.M.Smith, Eds.), pp.35-60. Oxford University Press, London.

[5] Berger, J. and Chen, M.-H. (1994), Predicting retirement patterns: prediction for a multinomial distribution with constrained parameter space. *The Statistician*, forthcoming.

[6] Bernardo, J.M. (1979). Reference posterior distributions for Bayes inference. *J. Roy. Statist. Soc. Ser.* **B 41** 113-147 (with discussion).

[7] Fraser, D.A.S., Monette, G. and Ng, K.W. (1984). Marginalization, likelihood, and structural models. *Multivariate Analysis VI*, (P.R. Krishnaiah, Ed.), pp.209-218. North-Holland, Amsterdam.

[8] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and related problems. *J. Amer. Statist. Assoc.* **72** 320-338.

[9] Jennrich, R.I. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42** 805-820.

[10] Laird, N., Lange, N. and Stram, A. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *J. Amer. Statist. Assoc.* **82** 97-105.

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

[12] Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons, New York.

[13] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[14] Swamy, P.A.V.B. (1971). *Statistical Inference in Random Coefficient Regression Models.* Springer-Verlag, New York.

[15] Tracy, D.S. and Jinadasa, K.G. (1988). Patterned matrix derivatives. *Canad. J. of Statist.* **16** 411-418.

[16] Wiens, D.P. (1985). On some pattern-reduction matrices which appear in statistics. *Linear Algebra Appl.* **67** 233-258.

[17] Yang, R. and Berger, J.O. (1993). Estimation of a covariance matrix using the reference prior. Technical Report **93-13C**, Department of Statistics, Purdue University.

[18] Ye, K.Y. (1991). Bayesian reference prior analysis on the ratio of one-way random effect model. Technical Report **91-22**, Department of Statistics, Virginia Polytechnic Institute and State University.

[19] Ye, K.Y. and Berger, J. (1991). Noninformative priors for inference in exponential regression models. *Biometrika* **78** 645-656.