2019

Bayesian analysis for the intraclass model and for the quantile semiparametric mixed-effects double regression models

Duo Zhang
Michigan Technological University, duoz@mtu.edu

BAYESIAN ANALYSIS FOR THE INTRACLASS MODEL AND FOR THE QUANTILE SEMIPARAMETRIC MIXED-EFFECTS DOUBLE REGRESSION MODELS

By

Duo Zhang

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2019

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Mathematical Sciences.

Department of Mathematical Sciences

Dissertation Advisor:	*Dr. Min Wang*
Committee Member:	*Dr. Yu Cai*
Committee Member:	*Dr. Jianping Dong*
Committee Member:	*Dr. Renfang Jiang*
Department Chair:	*Dr. Mark S. Gockenbach*

Contents

List of Figures

List of Tables

Preface

It has been 5 years since I entered the intriguing world of Bayesian Statistics in 2014. My research mainly focuses on Bayesian hypothesis testing, high-dimensional data analysis, and semiparametric and nonparametric models. The dissertation is submitted to fulfill the graduation requirements of Doctor of Philosophy at Michigan Technological University. The research was undertaken under the supervision of Dr. Min Wang in the department of Mathematical Sciences from September 2014 to May 2019. In my five years Ph.D. life, I have gained valuable academic knowledge and colorful life experience. Thanks to my university and the peaceful city Houghton, I had the most memorable period of my life.

Two of my dissertation-derived papers have been published by *Statistics & Probability Letters*, and *Statistical Theory and Related Fields*. I am very grateful to the editor, the associate editor, and anonymous referees of the two journals for their comments and suggestions that led to improvements of the papers.

I would like to express my heartfelt gratitude to my advisor Dr. Min Wang for his valuable guidance, encouragement, and patience throughout my five-year graduate studies and providing valuable suggestions in this research project. I wish to thank Drs. Yu Cai, Jianping Dong, and Renfang Jiang for their suggestive advice, help and serving on my dissertation committee. In addition, I would like to thank all of the faculties in my department, since without your cooperation and assistance I would not have been able to accomplish this dissertation. My parents Mrs. Yan Xu and Mr. Xingbin Zhang deserve a special note of thanks: your unconditional support and selfless love have always been the motivation of my progress.

Abstract

This dissertation consists of three distinct but related research projects. The first two projects focus on objective Bayesian hypothesis testing and estimation for the intraclass correlation coefficient in linear models. The third project deals with Bayesian quantile inference for the semiparametric mixed-effects double regression models.

In the first project, we derive the Bayes factors based on the divergence-based priors for testing the intraclass correlation coefficient (ICC). The hypothesis testing of the ICC is used to test the uncorrelatedness in multilevel modeling, and it has not well been studied from an objective Bayesian perspective. Simulation results show that the two sorts of Bayes factors have good performance in the hypothesis testing. Moreover, the Bayes factors can be easily implemented due to their unidimensional integral expressions.

In the second project, we consider objective Bayesian analysis for the ICC in the context of normal linear regression model. We first derive two objective priors for the unknown parameters and show that both result in proper posterior distributions. Within a Bayesian decision-theoretic framework, we then propose an objective Bayesian solution to the problems of hypothesis testing and point estimation of the ICC based on a combined use of the intrinsic discrepancy loss function and objective

priors. The proposed solution has an appealing invariance property under one-to-one reparameterization of the quantity of interest. Simulation studies are conducted to investigate the performance the proposed solution. Finally, a real data application is provided for illustrative purposes.

In the third project, we study Bayesian quantile regression for semiparametric mixed effects model, which includes both linear and nonlinear parts. We adopt the popular cubic spline functions for the nonlinear part and model the variance of the random effect as a function of the explanatory variables. An efficient Gibbs sampler with the Metropolis-Hastings algorithm is proposed to generate posterior samples of the unknown parameters from their posterior distributions. Simulation studies and a real data example are used to illustrate the performance of the proposed methodology.

Chapter 1

Introduction

This dissertation consists of three distinct but related research projects. The first two projects deal with objective Bayesian analysis for the intraclass correlation coefficient (ICC) of normal linear regression models. In the first project, we derive the Bayes factors based on the divergence-based priors for testing the presence of the ICC in linear models. In the second project, we study the problems of hypothesis testing and parameter estimation the ICC from a Bayesian decision-theoretic viewpoint. In the third project, we consider Bayesian quantile regression for the semiparametric mixed-effects models.

For illustrative purposes, we here briefly overview statistical inference for the ICC from a frequentist perspective. Donner [28] suggested the use of random effects one-way the analysis of variance (ANOVA) for making inference for the ICC. To be more specific, the one-way ANOVA with random effects is given by

$$y_{ij} = \mu + a_j + e_{ij}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, k, \tag{1.1}$$

where y_{ij} is the response i of subject j, μ is the grand mean of all the observations in the population, the treatment effects a_j are identically distributed with mean 0 and variance σ_a^2. Here, the error term $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, where $\overset{iid}{\sim}$ represents "independent and identically distributed". It can be shown that the variance of y_{ij} is given by $\sigma^2 = \sigma_a^2 + \sigma_e^2$. Then the ICC can be defined as

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}. \tag{1.2}$$

The unbiased estimates of σ_e^2 and σ_a^2 are given by

$$s_e^2 = \frac{SSE}{k(n-1)} \quad \text{and} \quad s_a^2 = \frac{SSTR}{n(k-1)} - \frac{SSE}{kn(n-1)},$$

respectively, where SSE$= \sum_{i=1} \sum_{j=1} (y_{ij} - \bar{y}_{i.})$ is the sum of squared errors of prediction and SSTR $= \sum_{i=1} n(y_{i.} - \bar{y}_{..})$ is the treatment sum of squares. Thus, the point estimate of ICC is $\hat{\rho} = s_a^2/(s_a^2 + s_e^2)$. It deserves mentioning that $\hat{\rho} = 0$ indicates no variation between groups ($s_a^2 = 0$) and that $\hat{\rho} = 1$ indicates no variation within groups ($s_e^2 = 0$). As commented by Box and Tiao [17], the classical unbiased estimates of s_a^2 can be a negative value even if the true value of σ_a^2 is nonnegative. This could be viewed a serious disadvantage of using these estimates within a frequentist framework; see, also, Wang and Sun [66].

The hypothesis testing problem of $H_1 : \rho = 0$ versus $H_2 : \rho \neq 0$ can be conducted by using the F-test statistic given by

$$F = \frac{\text{SSTR}}{k-1} / \frac{\text{SSE}}{k(n-1)},$$

which follows an F-distribution with degrees of freedom $k-1$ and $k(n-1)$ under the null hypothesis. For decision making at the α-th significance level, the null hypothesis is rejected if $F > F_{1-\alpha,k-1,k(n-1)}$, where $F_{1-\alpha,k-1,k(n-1)}$ is the $100\alpha\%$ upper percentage point of the F distribution with $(p-1)$ and $k(n-1)$ degrees of freedom.

In Chapter 2, we consider an objective Bayesian procedure for the hypothesis testing problem of the ICC in normal linear regression model. We derive the Bayes factors based on the divergence-based priors for testing the presence of the ICC. It turns out that the proposed Bayes factors only have unidimensional integral expressions and perform very well through numerous simulation studies.

In Chapter 3, we study the hypothesis testing and point estimation problems for the ICC from a decision-theoretical viewpoint. It is well-known that the choice of loss function plays a central role in the statistical decision theory. By adopting the intrinsic discrepancy as the loss function, we develop the Bayesian reference criterion for testing and estimating the ICC. The performance of the proposed approach is illustrated by simulation studies.

In Chapter 4, we propose the Bayesian quantile regression for the semiparametric mixed effects models. We employ the asymmetric Laplace distribution (ALD) for the error term. The convenient choice of ALD allows us to set the quantile in advance and the resulting posterior under a flat prior is the usual quantile regression estimates. We develop an efficient Gibbs sampler with the Metropolis-Hastings algorithm for the posterior sampling. The performance of the proposed procedure is examined through extensive simulation studies and a real-data application.

In Chapter 5, we discuss some future work based on the three projects. We consider developing the Bayes factor testing procedures based on the divergence-based priors in the network autocorrelation model. In addition, we plan to deal with Bayesian variable selection in the quantile semiparametric mixed effects models.

Chapter 2

Objective Bayesian Inference for the Intraclass Correlation Coefficient in Linear Models[1]

We outline objective Bayesian testing procedure for the intraclass correlation coefficient in linear models. For it, we derive the Bayes factors based on the divergence-based priors, which have unidimensional integral expressions and can thus be easily approximated numerically.

[1]This chapter has been published as an article in *Statistics & Probability Letters* (Zhang and Wang [73]). Reprinted with permission D.1.

2.1 Introduction

Consider the intraclass model of the form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, n, \tag{2.1}$$

where $\mathbf{y}_i$ is a $k \times 1$ ($k \geq 2$) vector of response variables, $\mathbf{X}_i$ is a $k \times p$ design matrix of $(p-1)$ regressors (assuming the first column is ones) with $p < k$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters. We assume that the random error $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_k, \sigma^2\mathbf{V})$, where $\overset{iid}{\sim}$ stands for "independent and identically distributed", $\mathbf{0}_k$ is a $k \times 1$ vector of zeros, and $\mathbf{V} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ with $\mathbf{I}_k$ being a $k \times k$ identity matrix and $\mathbf{J}_k$ being a $k \times k$ matrix containing only ones. The parameter ρ is often referred as the intraclass correlation coefficient (for short, ICC). It can be easily shown that $\rho \in (-(k-1)^{-1}, 1)$ is the necessary and sufficient condition for positive-definiteness of the covariance matrix $\mathbf{V}$. When $\rho = 0$, the intraclass model becomes the classical linear normal model with independent errors.

For notational simplicity, let

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}_{nk\times 1}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}_{nk\times p}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{nk\times 1}.$$

The above model can be represented in a rather compact form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.2}$$

where $\boldsymbol{\varepsilon}$ follows an nk-dimensional multivariate normal distribution with mean $\mathbf{0}_{nk}$ and covariance matrix $\sigma^2\mathbf{W}$, where $\mathbf{W} = \mathbf{I}_n \otimes \mathbf{V}$ with $\otimes$ being the Kronecker product. The probability density function (pdf) of $\mathbf{y}$ is given by

$$f(\mathbf{y} \mid \rho, \boldsymbol{\nu}) = (2\pi)^{-\frac{kn}{2}} \mid \sigma^2\mathbf{W}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{W}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \tag{2.3}$$

where $\boldsymbol{\nu} = (\sigma^2, \boldsymbol{\beta})$. We are interested in testing

$$H_1 : \rho = 0 \quad \text{versus} \quad H_2 : \rho \neq 0, \tag{2.4}$$

which can be equivalently expressed as the model selection problem of two competing models

$$M_1 : f_1(\mathbf{y} \mid \boldsymbol{\nu}) = f(\mathbf{y} \mid 0, \boldsymbol{\nu}) \quad \text{versus} \quad M_2 : f_2(\mathbf{y} \mid \rho, \boldsymbol{\nu}) = f(\mathbf{y} \mid \rho, \boldsymbol{\nu}). \tag{2.5}$$

The ICC has a lengthy history of practical applications in various fields of study as a coefficient of reliability, such as epidemiologic research, genetics, psychology, and sociology; see, for example, Barkto [4], Lin et al. [57], to name just a few. One practical example is, in the multilevel modeling, ICC is often adopted to measure the strength of correlation in a hierarchical data, which helps researchers determine if the uncorrelatedness assumption is violated in the data. Another practical example is the following, extracted from Chapter 5.2 of Frees [33]: twenty-seven individuals including 16 boys and 11 girls were measured for distances from the pituitary to the pteryomaxillary fissure in millimeters, at ages 8, 10, 12, and 14. In this case, the distance y_{ij} measured in millimeters is the response for individual i measured at age j, the design matrix consists of two columns with the first being age and the second being gender (1 for males and 0 for females), and $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_4, \sigma^2\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_4+\rho\mathbf{J}_4$. We are interested in studying how strong the individuals resemble each other (i.e., $\rho = 0$, where ρ represents the resemblance among individuals).

Bayesian estimation of ρ has been conducted in the literature. Ghosh and Heo [37] considered Bayesian credible intervals for ρ based on objective priors, whereas they did not study the hypothesis testing of ρ. Later on, Lee and Kim [55] studied the

Bayesian Reference Criterion (BRC) for making inference of ρ, whereas the BRC depends on an arbitrary threshold when making a formal decision. To the best of my knowledge, the hypothesis testing of ρ in (2.4) has not well been studied from an objective Bayesian perspective.

We develop an objective Bayesian solution to compare two competing models in (2.5), see Berger and Pericchi [8] for a nice discussion about the advantages of using Bayesian methods for model comparison. A natural way for comparing two competing models is the Bayes factor (Kass and Raftery [45]), which has an intuitive meaning of "measure of evidence" in favor of a model under the hypotheses. The Bayes factor (BF) in favor of M_2 and against M_1 is defined as

$$\mathrm{BF}_{21} = \frac{p(\mathbf{y} \mid M_2)}{p(\mathbf{y} \mid M_1)} = \frac{\int f_2(\mathbf{y} \mid \rho, \boldsymbol{\nu})\pi_2(\rho, \boldsymbol{\nu})\, d\rho\, d\boldsymbol{\nu}}{\int f_1(\mathbf{y} \mid \boldsymbol{\nu})\pi_1(\boldsymbol{\nu})\, d\boldsymbol{\nu}}, \tag{2.6}$$

where $\pi_1(\boldsymbol{\nu})$ and $\pi_2(\rho, \boldsymbol{\nu})$ are the prior probabilities under models M_1 and M_2, respectively. In general, when $\mathrm{BF}_{21} > (<) 1$, it indicates the data are more likely to have occurred under M_2 (M_1). For instance, $\mathrm{BF}_{21} = 5$ indicates that the data are 5 times more likely under M_2 than under M_1 ($\mathrm{BF}_{12} = 1/\mathrm{BF}_{21} = .2$). The posterior probability of M_1 given the data can be represented as

$$p(M_1 \mid \mathbf{y}) = \left[1 + \mathrm{BF}_{21}\frac{p(M_2)}{p(M_1)}\right]^{-1},$$

where $p(M_2)/p(M_1)$ is the prior model odds between two models or hypotheses, which is assumed to be 1 in this paper. Unlike the frequentist P-value test, the value of $p(M_1 \mid \mathbf{y})$ (or $p(M_2 \mid \mathbf{y}) = 1 - p(M_1 \mid \mathbf{y})$) allows practitioners to quantify the support in a probability scale that the data provide for one hypothesis over another.

A critical ingredient of deriving the BF is to specify priors for the unknown parameters under hypotheses. In the absence of prior knowledge, noninformative priors are usually preferred, such as the Jeffreys prior (Jeffreys [43]) and the reference prior (Bernardo [9]), whereas these priors are often improper and result in the BF up to an undefined multiplicative constant. Bayarri and García-Donato [5] proposed an attractive way to obtain noninformative while proper priors, (so-called the divergence-based (DB) priors). Since then, they have been implemented in practical applications. For instance, García-Donato and Sun [34] adopted the DB priors for testing of no difference between groups in the one-way random-effects model. Kim et al. [47] considered the DB priors for testing the autocorrelation coefficient in linear models with first-order autoregressive residuals. Note that derivation of these priors needs the parameters to be orthogonal (Kass and Vaidyanathan [46]) or at least approximately orthogonal for moderate or large sample sizes (García-Donato and Sun [34]).

This paper derives the BF associated with the DB priors for comparing two competing models in (2.5). The resulting Bayes factors have unidimensional integral expressions that can be numerically approximated in most statistical software, such as R and

SAS. Numerical results show that they perform very well in terms of the sum of the frequentist type I and type II error probabilities, i.e., the probability of incorrectly choosing H_2 while H_1 is true and the probability of incorrectly choosing H_1 while H_2 is true, respectively.

The remainder of this paper is organized as follows. In Section 2.2, we derive the DB priors and their resulting BFs. In Section 2.3, we conduct simulations to evaluate the performance of the BFs. Some concluding remarks are provided in Section 2.4, with additional proofs given in the Appendix A.

2.2 The DB priors and the resulting BFs

In this section, we provide an objective Bayesian solution for the problem of hypothesis testing in (2.4) based on the DB priors. In Subsection 2.2.1, we consider objective priors for ρ after an orthogonal reparameterization (Cox and Reid [23]) of $(\rho, \boldsymbol{\beta}, \sigma^2)$ of model in (2.1). We derive the conditional sum-DB and min-DB priors for ρ (Subsection 2.2.2) and obtain the BFs associated with these priors (Subsection 2.2.3).

2.2.1 Objective priors for the unknown parameters

As mentioned in Section 2.1, derivation of the DB priors require the parameters to be orthogonal. With orthogonal parameters, the off-diagonal elements of the Fisher information matrix are all 0. Thus, orhtogonality is an important simplification for derivation of the noninformative priors. First we find an orthogonal transformation of $(\rho, \boldsymbol{\beta}, \sigma^2)$, then we can employ the noninformative priors. For model in (2.2), we follow the orthogonal transformation of the Fisher information matrix (Ghosh and Heo [37]) and let

$$\theta_1 = \rho, \quad \theta_2 = \frac{1}{\sigma^2}(1-\rho)^{-(k-1)/k}\big(1+(k-1)\rho\big)^{-1/k}, \quad \boldsymbol{\theta}_3 = \boldsymbol{\beta}.$$

Under this orthogonal transformation, θ_1 is orthogonal to $\boldsymbol{\theta}_2$ and θ_3; thus, $\boldsymbol{\nu} = (\boldsymbol{\theta}_2, \theta_3)$ can be viewed as common parameters of both models in (2.3). The pdf of $\mathbf{y}$ in (2.2) can be re-expressed as

$$f(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) \propto \mid \theta_2^{-1}\boldsymbol{\Sigma}\mid^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)^T(\theta_2^{-1}\boldsymbol{\Sigma})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)\Big\},$$

where $\boldsymbol{\Sigma} = (1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k}\mathbf{W}$, $\mathbf{W} = \mathbf{I}_n \otimes \mathbf{V}$ and $\mathbf{V} = (1-\theta_1)\mathbf{I}_k + \theta_1\mathbf{J}_k$. The hypothesis testing problem of ρ in (2.4) is equivalent to $H_1 : \theta_1 = \theta_0 \quad \text{versus} \quad H_2 : \theta_1 \neq \theta_0$ with $\theta_0 = 0$, and thus it becomes the problem of comparing two competing models

$$M_1 : f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3) = N_{nk}\big(\mathbf{X}\boldsymbol{\theta}_3, \theta_2^{-1}\mathbf{I}_{nk}\big),$$
$$M_2 : f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) = N_{nk}\big(\mathbf{X}\boldsymbol{\theta}_3, \theta_2^{-1}\boldsymbol{\Sigma}\big). \tag{2.7}$$

This paper adopts the second-order matching prior (Datta and Mukerjee [26]) for $(\theta_1, \theta_2, \boldsymbol{\theta}_3)$ to develop the DB priors, because of its nice frequentist coverage probability studied by Ghosh and Heo [37]. This prior under the above model M_2 is given by

$$\pi^N(\theta_1, \theta_2, \boldsymbol{\theta}_3) \propto (1-\theta_1)^{-1}\big(1+(k-1)\theta_1\big)^{-1}\theta_2^{-1}. \tag{2.8}$$

By following [46], we assume $(\theta_2, \boldsymbol{\theta}_3)$ to have the same meaning to both models and thus specify a common (even improper) prior given by $\pi^N(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1}$. Based on the prior in (2.8), we define a noninformative prior for θ_1 (the parameter of interest) as

$$\pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) \propto (1-\theta_1)^{-1}\big(1+(k-1)\theta_1\big)^{-1}. \tag{2.9}$$

This leads to the following noninformative priors for the unknown parameters of models in (2.7)

$$\pi^N(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1},$$
$$\pi^N(\theta_1, \theta_2, \boldsymbol{\theta}_3) \propto \pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)\pi^N(\theta_2, \boldsymbol{\theta}_3),$$

which will be used to derive the DB priors discussed in the following section.

2.2.2 The DB priors

The DB priors, proposed by Bayarri and García-Donato [5], are designed to use other formal rules to construct objective priors of the new parameters under the alternative hypothesis. They have shown that these priors are a generalization of Jeffreys-Zillow-Siow priors for the model selection problems in linear models and are quite suitable for Bayesian hypothesis testing under certain scenarios in which other proposals may fail. Note that the DB measures are derived based on the measure of the direct Kullback-Leibler (KL) divergence of the models under comparison, raised to a negative power. The KL divergence between M_1 and M_2 in (2.7) is given by

$$\mathrm{KL}[\theta_0 : \theta_1] = \int \log \frac{f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)}{f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)} f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) d\mathbf{y}. \tag{2.10}$$

We usually take the sums or minimum of the KL directed divergences to obtain the symmetry property. Then the sum-DB measure is given by

$$D^S[\theta_0, \theta_1] = \mathrm{KL}[\theta_0 : \theta_1] + \mathrm{KL}[\theta_1 : \theta_0], \tag{2.11}$$

and the min-DB measure is given by

$$D^M[\theta_0,\theta_1] = 2 \times \min\{\mathrm{KL}[\theta_0 : \theta_1],\ \mathrm{KL}[\theta_1 : \theta_0]\}. \tag{2.12}$$

For our testing problem at hand, we follow Definition 2.2.1 of Bayarri and García-Donato [5] and obtain both the sum and minimum DB priors summarized in the following proposition with proofs given in the Appendix A.1.

Definition 2.2.1 *Let*

$$c(q) = \int_{\theta_1} (1+\overline{D}[\theta_0,\theta_1])^{-q}\ \pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) d\theta_1$$

and

$$\underline{q} = \inf\{q \geq 0 : c(q) < \infty\}, \quad q_* = \underline{q} + 2^{-1},$$

where $\overline{D}[\theta_0,\theta_1] = D[\theta_0,\theta_1]/n^*$ *is the mean divergence measure and* $\pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) = (1-\theta_1)^{-1}\big(1+(k-1)\theta_1\big)^{-1}$ *is the conditional prior for* θ_1 *from (2.9).* n^* *is the effective sample size equal to the number of data points. Then the DB prior under* M_1 *is* $\pi^D(\theta_2, \boldsymbol{\theta}_3) = \pi^N(\theta_2, \boldsymbol{\theta}_3) = \theta_2^{-1}$ *and under* M_2 *is* $\pi^D(\theta_1, \theta_2, \boldsymbol{\theta}_3) = \pi^D(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)\pi^N(\theta_2, \boldsymbol{\theta}_3)$*, where the conditional DB prior for* θ_1 *is given by*

$$\pi^D(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) = c^{-1}(q_*)(1+\overline{D}[\theta_0,\theta_1])^{-q_*}\ \pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3).$$

The conditional sum-DB prior π^S and the conditional min-DB prior $\pi^M(\cdot)$ are defined when D is chosen to be $D^S(\cdot)$ in (2.11) or $D^M(\cdot)$ in (2.12) respectively. In what follows, we refer to their corresponding c's and q's as c_S, $\underline{q}^S$, q_*^S and c_M, $\underline{q}^M$, q_*^M, respectively.

Proposition 1 *Under* M_2 *in (2.7), the conditional sum-DB prior for θ_1 is*

$$\pi^S(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) = \frac{\sqrt{2}}{c_S}\left[\frac{(1-\theta_1)^{1/k+1}}{(1+(k-1)\theta_1)^{1/k-2}} + \frac{(1-\theta_1)^{-1/k+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-1}}\right]^{-\frac{1}{2}}, \tag{2.13}$$

where

$$c_S = \sqrt{2}\int_{-\frac{1}{k-1}}^{1}\left[\frac{(1-\theta_1)^{1/k+1}}{(1+(k-1)\theta_1)^{1/k-2}} + \frac{(1-\theta_1)^{-1/k+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-1}}\right]^{-\frac{1}{2}} d\theta_1.$$

The conditional min-DB prior for θ_1 is

$$\begin{aligned}\pi^M(\theta \mid \theta_2, \boldsymbol{\theta}_3) = c_M^{-1}\Big[&(1-\theta_1)^{\frac{1-2k}{2k}}\big(1+(k-1)\theta_1\big)^{-\frac{k+1}{2k}}\big(1+(k-2)\theta_1\big)^{-\frac{1}{2}} I(\theta_1 \geq 0) \\ &+ (1-\theta_1)^{-\frac{k+1}{2k}}\big(1+(k-1)\theta_1\big)^{\frac{1-2k}{2k}} I(\theta_1 < 0)\Big] \end{aligned} \tag{2.14}$$

where $I(A)$ is an indicator function, such that $I(A) = 1$ if A is true and 0 otherwise, *and*

$$c_M = \int_{-\frac{1}{k-1}}^{1} \Big[(1-\theta_1)^{\frac{1-2k}{2k}} \big(1+(k-1)\theta_1\big)^{-\frac{k+1}{2k}} \big(1+(k-2)\theta_1\big)^{-\frac{1}{2}} I(\theta_1 \geq 0)$$
$$+ (1-\theta_1)^{-\frac{k+1}{2k}} \big(1+(k-1)\theta_1\big)^{\frac{1-2k}{2k}} I(\theta_1 < 0)\Big] d\theta_1.$$

It deserves mentioning that both conditional sum-DB and min-DB priors are proper and thus can be used for the new parameter θ_1 under the alternative hypothesis. Thus, the BFs based on these priors are also well-defined for comparing two competing models in (2.7) discussed in the following subsection.

2.2.3 The Bayes factors based on the DB priors

We derive the BFs for comparing two competing models in (2.7) based on the priors in Subsections 2.2.1 and 2.2.2, which are given by

$$\pi^D(\theta_2, \boldsymbol{\theta}_3) \propto \frac{1}{\theta_2},$$
$$\pi^D(\theta_1, \theta_2, \boldsymbol{\theta}_3) \propto \pi^D(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)\pi^D(\theta_2, \boldsymbol{\theta}_3),$$

where $\pi^D(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)$ is the conditional DB prior for θ_1 in Proposition 1 with $D = S$ and M. The resulting BF is defined as

$$\mathrm{BF}_{21} = \frac{\iiint f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)\pi^D(\theta_1, \theta_2, \boldsymbol{\theta}_3)\, d\theta_1\, d\theta_2\, d\boldsymbol{\theta}_3}{\iint f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)\pi^D(\theta_2, \boldsymbol{\theta}_3)\, d\theta_2\, d\boldsymbol{\theta}_3}. \tag{2.15}$$

We summarize the resulting BFs in the following theorem with proofs provided in the Appendix A.2.

Theorem 1 *The BF associated with the conditional sum-DB prior in (2.13) in favor of M_2 in (2.7) is given by*

$$\mathrm{BF}_{21}^S(\mathbf{Y}) = c_S^{-1}|\mathbf{X}^T\mathbf{X}|^{\frac{1}{2}}\left[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\right]^{\frac{nk-p}{2}} \int_{-\frac{1}{k-1}}^{1} h_S(\theta_1)d\theta_1, \tag{2.16}$$

where $\mathbf{H}_1 = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, $\mathbf{H}_2 = \mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}$ *with* $\mathbf{W} = \mathbf{I}_n \otimes \mathbf{V}$, c_S *is a constant defined in Proposition 1, and*

$$\begin{aligned} h_S(\theta_1) =& \sqrt{2}\left[\frac{(1-\theta_1)^{1/k+n(k-1)+1}}{(1+(k-1)\theta_1)^{1/k-n-2}} + \frac{(1-\theta_1)^{-1/k+n(k-1)+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-n-1}}\right]^{-\frac{1}{2}} \\ &\times |\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}|^{-\frac{1}{2}}\left[\mathbf{y}^T(\mathbf{W}^{-1} - \mathbf{H}_2)\mathbf{y}\right]^{-\frac{nk-p}{2}}. \end{aligned}$$

The BF associated with the conditional min-DB prior in (2.14) in favor of M_2 in

(2.7) is given by

$$\mathrm{BF}_{21}^{M}(\mathbf{Y}) = c_M^{-1}|\mathbf{X}^T\mathbf{X}|^{\frac{1}{2}}\left[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\right]^{\frac{nk-p}{2}} \int_{-\frac{1}{k-1}}^{1} h_M(\theta_1)d\theta_1, \tag{2.17}$$

where c_M is a constant defined in Proposition 1 and

$$\begin{aligned} h_M(\theta_1) = &\Big[(1-\theta_1)^{\frac{1}{2k}+\frac{n-nk-2}{2}}\big(1+(k-1)\theta_1\big)^{-\frac{1}{2k}-\frac{n+1}{2}}\big(1+(k-2)\theta_1\big)^{-\frac{1}{2}}I(\theta_1 \geq 0) \\ &+ (1-\theta_1)^{-\frac{1}{2k}+\frac{n-nk-1}{2}}\big(1+(k-1)\theta_1\big)^{\frac{1}{2k}-\frac{n+2}{2}}I(\theta_1 < 0)\Big] \\ &\times |\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}|^{-\frac{1}{2}}\left[\mathbf{y}^T(\mathbf{W}^{-1} - \mathbf{H}_2)\mathbf{y}\right]^{-\frac{nk-p}{2}}. \end{aligned}$$

Note that the Bayes factors in (2.16) and (2.17) have really simple expressions with a unidimensional integral and can thus be numerically approximated by using standard statistical package, such as R and SAS.

2.3 Simulation study

In this section, we undertake simulation studies to investigate the performance of two Bayes factors (BF_{21}^S and BF_{21}^M) for testing the hypotheses in (2.4). Without loss of generality, we choose $\sigma^2 = 1$, $\boldsymbol{\beta} = (1, 0.5, -1)^T$, and $\mathbf{X}_1 = \cdots = \mathbf{X}_n = \mathbf{X}_c$, where $\mathbf{X}_c$ is a $k \times 3$ matrix generated from a uniform distribution over $(-2, 2)$ with mean 0. Since $\mathbf{X}_c$ is already centered, we do not need an intercept column. We then generate data $\mathbf{y}_i$ from the same multivariate normal distribution $\mathrm{N}_k(\mathbf{X}_c\boldsymbol{\beta}, \sigma^2\mathbf{V})$, where $\mathbf{V} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ with ρ ranging from $-(k-1)^{-1} + 0.01$ to 0.99 in increment of 0.01. To assess the influence of the sample size and dimension of random variables, we take $n = 10$, 20, 50 and $k = 3$, 6, 9. For each case, we simulate $N = 5000$ data sets with various choices of n and k. The decision criterion used in this paper is to choose H_2 if the value of the BF exceeds 1 and H_1 otherwise (Kass and Raftery [45]).

The relative frequencies of rejection of H_1 under different combinations of n and k are depicted in Figure 2.1. For further illustrative purposes, we also report the relative frequency of rejecting H_1 and the average posterior probability of H_2 for certain chosen values of n, k, and ρ in Table 2.1. Rather than providing exhaustive results based on these simulations, we merely highlight the most important findings as follows:

(i) Two BFs perform well under across all simulation scenarios. When H_1 is false (i.e., $\rho \neq 0$), BF_{21}^S slightly outperforms BF_{21}^M in terms of the relative frequency of rejecting H_1 and the average posterior probability of H_2, whereas they behave similarly when n becomes large.

(ii) When k is moderate or large, the BF_{21}^S outperforms BF_{21}^M, because the latter is generally more in favor of H_1 than the former, especially when H_1 true, leading to its worse performance than BF_{21}^S when H_1 is wrong. The amount of differences between them disappears quickly with an increasing value of either $|\rho|$ or n.

(iii) As expected, the relative frequency of rejection of H_1 will significantly increases as $|\rho|$ increases. Also, as either n or k becomes large, the relative frequency of rejection of H_1 also increases when H_1 is false, This is mainly because in statistical theory the Bayesian procedures generally have better performance with an increasing value of n.

(iv) Simulation studies suggest that the BF_{21}^{S} should be preferred in practical applications, because the sum of the type I and type II error probabilities of BF_{21}^{S} is smaller than the one of BF_{21}^{M} across all the considered simulations.

We now investigate the frequentist coverage probability of the marginal posteriors of ρ under these two priors. Let α be the left tail probability and $\rho^{(\alpha)}$ be the corresponding quantile of the posterior distribution $\pi(\rho \mid \text{Data})$ under the DB priors π^S or π_M. Theoretically, it follows $F(\rho^{(\alpha)}) = \int_{-\infty}^{\rho^{(\alpha)}} \pi(\rho \mid \text{Data}) d\rho = \alpha$. Letting $P(\alpha \mid \rho) = P(F(\rho) < \alpha \ \mid \rho, \text{ Data}) = P(\int_{-\infty}^{\rho} \pi(\rho \mid \text{Data}) d\rho < \alpha \ \mid \rho, \text{ Data})$, we observe that $P(\alpha \mid \rho)$ should be very close to α if the chosen prior performs well with respect to the probability matching criterion. The last two columns of Table 2.1 shows the estimated 95% coverage of the posterior distributions between two priors under different scenarios, which is denoted by $P(95\% \mid \pi^S \text{ or } \pi^M) = P(97.5\% \mid \rho) - P(2.5\% \mid \rho)$. We observe that estimated 95% coverage of the posterior distribution under different priors is very close to the frequentist coverage probabilities even for small sample sizes.

2.4 Concluding remarks

We derived the DB priors (Bayarri and García-Donato [5]) and their resulting BFs for the intraclass correlation coefficient in linear models, which are shown to have unidimensional integral expressions that can be easily implemented by practitioners. It deserves mentioning that the classical balanced one-way random effect model is just a special case of the intraclass model by letting $\sigma^2 = \sigma_a^2 + \sigma_e^2$ and $\rho = \sigma_a^2/\sigma_e^2 \in (0, 1)$, where σ_a^2 and σ_e^2 stand for the treatment and error variances, respectively; see García-Donato and Sun [34]. This observation motivates a study of generalizing our results to the unbalanced case with different number of observations in each group, which is currently under investigation and will be reported elsewhere.

Table 2.1

Relative frequencies of rejection of $H_1: \rho = 0$ (for short, $\mathrm{RF}(H_1)$) and the average posterior probabilities (square root of the MSE) of H_2 based on the BFs under the sum-DB and min-DB priors, respectively. Frequentist coverage of the 95% credible interval.

k	ρ	n	$\mathrm{RF}^S(H_1)$	$p^S(M_2 \mid \mathbf{y})$	$\mathrm{RF}^M(H_1)$	$p^M(M_2 \mid \mathbf{y})$	$P(95\% \mid \pi^S)$	$P(95\% \mid \pi^M)$
3	−0.1	10	0.1070	0.2757 (0.1700)	0.1000	0.2647 (0.1689)	95.40%	95.34%
		20	0.0778	0.2081 (0.1760)	0.0655	0.1871 (0.1703)	95.30%	95.28%
		50	0.1248	0.2254 (0.2275)	0.1142	0.2062 (0.2219)	94.66%	94.66%
	0.0	10	0.0672	0.2458 (0.1393)	0.0625	0.2358 (0.1387)	95.70%	95.66%
		20	0.0328	0.1587 (0.1183)	0.0258	0.1404 (0.1125)	94.98%	95.00%
		50	0.0185	0.1119 (0.1048)	0.0155	0.0985 (0.0987)	94.60%	94.60%
	0.2	10	0.2002	0.3353 (0.2164)	0.1930	0.3255 (0.2175)	95.06%	95.02%
		20	0.2395	0.3320 (0.2713)	0.2202	0.3093 (0.2697)	95.32%	95.30%
		50	0.4700	0.5036 (0.3325)	0.4438	0.4815 (0.3355)	95.12%	95.14%
	0.4	10	0.5128	0.5464 (0.2882)	0.5025	0.5378 (0.2914)	95.22%	95.22%
		20	0.7352	0.7131 (0.2956)	0.7163	0.6949 (0.3051)	95.04%	95.06%
		50	0.9768	0.9603 (0.1209)	0.9745	0.9564 (0.1285)	94.50%	94.52%
6	−0.1	10	0.2622	0.3740 (0.2427)	0.1675	0.2767 (0.2299)	96.10%	96.06%
		20	0.4580	0.4943 (0.2911)	0.3422	0.3989 (0.2941)	95.80%	95.76%
		50	0.8630	0.8136 (0.2355)	0.7910	0.7523 (0.2760)	94.86%	94.86%
	0.0	10	0.0503	0.2004 (0.1352)	0.0260	0.1294 (0.1137)	95.94%	95.92%
		20	0.0292	0.1490 (0.1192)	0.0155	0.0941 (0.0987)	95.42%	95.44%
		50	0.0195	0.1041 (0.1073)	0.0098	0.0649 (0.0881)	94.76%	94.76%
	0.2	10	0.4418	0.4970 (0.3156)	0.3598	0.4142 (0.3266)	94.78%	94.78%
		20	0.6882	0.6827 (0.3209)	0.6070	0.6159 (0.3489)	94.78%	94.76%
		50	0.9610	0.9459 (0.1517)	0.9438	0.9276 (0.1820)	95.04%	95.00%
	0.4	10	0.8655	0.8485 (0.2498)	0.8250	0.8091 (0.2875)	94.46%	94.64%
		20	0.9872	0.9789 (0.0942)	0.9810	0.9713 (0.1149)	94.88%	94.94%
		50	0.9998	0.9999 (0.0092)	0.9998	0.9998 (0.0113)	95.02%	95.08%
9	−0.1	10	0.9062	0.8141 (0.1930)	0.7622	0.6969 (0.2506)	95.38%	95.16%
		20	0.9980	0.9794 (0.0541)	0.9908	0.9585 (0.0912)	95.78%	95.70%
		50	1.0000	1.0000 (0.0001)	1.0000	1.0000 (0.0003)	95.49%	95.81%
	0.0	10	0.0455	0.1895 (0.1321)	0.0162	0.0981 (0.0993)	95.36%	95.34%
		20	0.0238	0.1424 (0.1129)	0.0078	0.0711 (0.0815)	95.00%	94.96%
		50	0.0160	0.0989 (0.1025)	0.0080	0.0488 (0.0745)	95.36%	95.36%
	0.2	10	0.6740	0.6778 (0.3221)	0.5648	0.5808 (0.3627)	94.96%	94.96%
		20	0.8952	0.8742 (0.2302)	0.8348	0.8211 (0.2839)	94.72%	94.74%
		50	0.9978	0.9951 (0.0426)	0.9942	0.9920 (0.0603)	94.86%	94.84%
	0.4	10	0.9632	0.9491 (0.1528)	0.9370	0.9255 (0.1957)	94.58%	94.58%
		20	0.9998	0.9985 (0.0172)	0.9990	0.9972 (0.0285)	94.72%	94.62%
		50	1.0000	1.0000 (0.0000)	1.0000	1.0000 (0.0001)	95.16%	95.22%

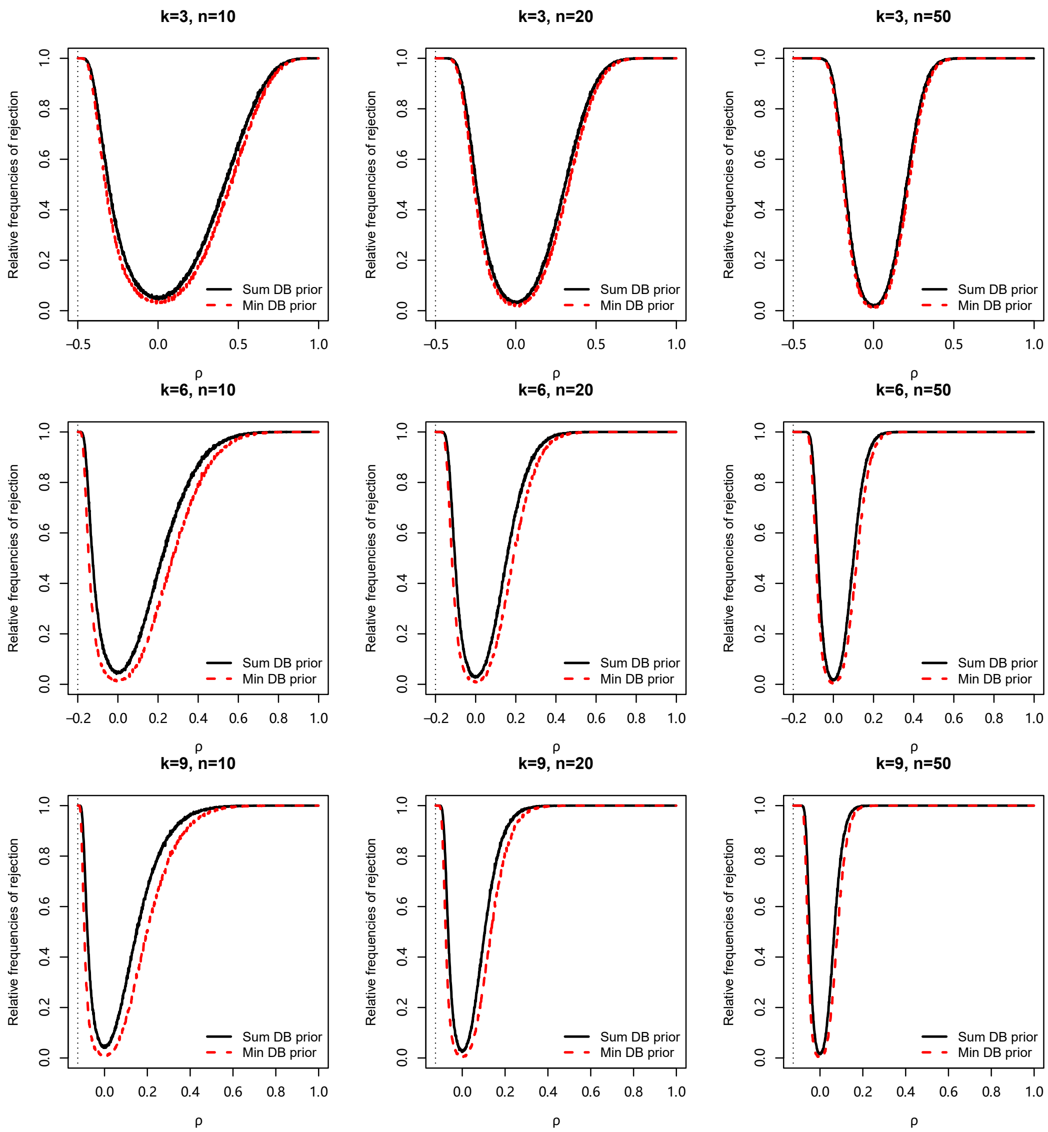


Figure 2.1: Relative frequencies of rejection of $H_1 : \rho = 0$ based on the BFs associated with the sum-DB and min-DB priors, respectively.

Chapter 3

Objective Bayesian Hypothesis Testing and Estimation for the Intraclass Model[1]

The intraclass correlation coefficient (ICC) plays an important role in various fields of study as a coefficient of reliability. In this paper, we consider objective Bayesian analysis for the ICC in the context of normal linear regression model. We first derive two objective priors for the unknown parameters and show that both result in proper posterior distributions. Within a Bayesian decision-theoretic framework, we then propose an objective Bayesian solution to the problems of hypothesis testing and

[1]This chapter has been published as an article in *Statistical Theory and Related Fields* (Zhang et al. [72]). Reprinted with permission D.2.

point estimation of the ICC based on a combined use of the intrinsic discrepancy loss function and objective priors. The proposed solution has an appealing invariance property under one-to-one reparametrization of the quantity of interest. Simulation studies are conducted to investigate the performance the proposed solution. Finally, a real-data application is provided for illustrative purposes.

3.1 Introduction

Consider the intraclass model of the form

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, n, \tag{3.1}$$

where $\mathbf{Y}_i$ is a $k \times 1$ vector of response variables, $\mathbf{X}_i$ is a $k \times p$ design matrix of $(p-1)$ regressors (assuming the first column is ones) and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown common regression coefficients. We assume that the random error $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_k, \sigma^2\boldsymbol{\Sigma})$, where $\overset{iid}{\sim}$ stands for "independent and identically distributed," $\mathbf{0}_k$ is a $k \times 1$ vector of zeros, and $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ with $\mathbf{I}_k$ being a $k \times k$ identity matrix and $\mathbf{J}_k$ being a $k \times k$ matrix containing only ones. The parameter ρ is often referred as the intraclass correlation coefficient (ICC). Note that $\rho \in (-(k-1)^{-1}, 1)$ is the necessary and sufficient condition for positive-definiteness of $\boldsymbol{\Sigma}$. When ρ is equal to 0, the intraclass model becomes the classical linear normal model with independent errors.

The ICC has been widely applied in various fields of study as a coefficient of reliability, from epidemiologic research to genetic studies; see, for example, Barkto [4], Fleiss [32], Lin et al. [57], among others. The analysis of the ICC transitionally consists of two branches, hypothesis testing and point estimation, and it has received attentions from two main statistical streams of thought: frequentists and Bayesians. From a frequentist viewpoint, Paul [59] considered the maximum likelihood estimate (MLE) of the ICC in a generalized model setting by solving iteratively a single estimating equation. Paul [60] developed the score tests for testing the significance of the interclass correlation in familial data. For Bayesian methods, Jelenkowska [44] studied Bayesian estimation of the ICC in the linear mixed model. Chung and Dey [22] considered Bayesian analysis of the ICC using the reference prior under a balanced variance components model. Later on, Ghosh and Heo [37] considered Bayesian credible intervals for ρ based on different objective priors and made comparisons among these priors in terms of matching the corresponding frequentist coverage probabilities.

It deserves mentioning that the problems of hypothesis testing and point estimation for ρ have not yet been studied within a decision-theoretical viewpoint. This motivates us to propose an objective Bayesian solution to these problems based on the Bayesian reference criterion (for short, BRC) (Bernardo and Rueda [15]). The proposed solution allows the researchers to simultaneously study important inference summaries of the ICC, including point estimation, credible interval estimation, and precise hypotheses. In addition, it enjoys various appealing properties: (i) it is invariant under one-to-one reparametrization of the parameter of interest ρ; (ii) it depends only on the assumed model, appropriate objective priors, and the observed data; (iii) it is appropriate to perform the hypothesis test: $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$ for any $\rho_0 \in (-(k-1)^{-1}, 1)$, and (iv) it can be easily approximated numerically in most statistical software and can thus be implemented by the practitioners from different fields.

The remainder of the paper is organized as follows. In Section 3.2, we derive two objective priors of the unknown parameters and discuss the propriety of their corresponding posterior distributions. In Section 3.3, we propose an objective Bayesian solution to both hypothesis testing and estimation problems of ρ from a decision-theoretical viewpoint. Section 3.4 investigates the performance of the proposed solution through simulations and a real data application. Some concluding remarks are provided in Section 3.5, with additional proofs given in the Appendix B.

3.2 Posterior distribution

For notational convenience, let $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ be $nk \times 1$ vectors and $\mathbf{X}$ is an $nk \times p$ design matrix, and they are given by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix},$$

respectively. The model in (3.1) can be expressed in a more compact way as

$$\mathbf{Y} \;=\; \mathbf{X}\boldsymbol{\beta} \;+\; \boldsymbol{\varepsilon}, \tag{3.2}$$

where $\boldsymbol{\varepsilon}$ follows an nk-dimensional normal distribution with mean vector $\mathbf{0}_{nk}$ and covariance matrix $\sigma^2\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$ is an nk-dimensional matrix and $\otimes$ denotes the Kronecker product. The likelihood function of the intraclass model in (3.2) is given by

$$\begin{aligned} p(\mathbf{Y} \mid \boldsymbol{\beta}, \sigma^2, \rho) \propto & |\sigma^2\boldsymbol{\Phi}|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Phi}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\} \\ \propto & (\sigma^2)^{-nk}(1-\rho)^{-n(k-1)/2}(1+(k-1)\rho)^{-n/2} \\ & \times \exp\left\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Phi}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\}, \end{aligned}$$

where $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$.

Bayesian analysis begins with prior specification for all the unknown parameters in the model. In the absence of relevant prior knowledge for $(\boldsymbol{\beta}, \sigma^2, \rho)$ in the above model, noninformative priors are often preferred. One of the most popular noninformative priors is the Jeffreys prior, which is proportional to the square root of the determinant of the Fisher information matrix. It can be shown that the Jeffreys prior is given by

$$\pi_J(\rho, \sigma^2, \boldsymbol{\beta}) \propto \left(\sigma^2\right)^{-(p+2)/2}(1-\rho)^{-1}(1+(k-1)\rho)^{-1}|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{1/2}. \tag{3.3}$$

Given that the parameter of interest is ρ, we integrate out $\boldsymbol{\beta}$ and σ^2 (i.e., $\pi_J(\rho \mid D) \propto \int\int f(\mathbf{Y} \mid \boldsymbol{\beta}, \sigma^2, \rho)\pi_J(\rho, \sigma^2, \boldsymbol{\beta})\, d\boldsymbol{\beta}\, d\sigma^2$) and obtain the marginal posterior density for ρ, denoted by $\pi_J(\rho \mid D)$, where D represents the observable data. It follows that

$$\pi_J(\rho \mid D) \propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1}\mathbf{S}(\rho)^{-nk/2}, \tag{3.4}$$

where $\mathbf{S}(\rho) = \mathbf{Y}'\big(\mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1}\big)\mathbf{Y}$. Note that when $\mathbf{X}_1 = \cdots = \mathbf{X}_n$, the prior in (3.4) can be simplified by replacing $\mathbf{S}(\rho)$ with $(\mathbf{Y} - \mathbf{X}'\hat{\boldsymbol{\beta}})'\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}'\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1}\bar{\mathbf{Y}}$ and $\bar{\mathbf{Y}} = \sum_{i=1}^{n}\mathbf{Y}_i/n$. The simplified version is just the Jeffreys prior derived by Ghosh and Heo [37].

One may argue that, when we aim at a subset of the parameters with the rest treated as nuisance parameters, the direct use of the Jeffreys prior may sometimes be unsatisfactory. To overcome such a pitfall, Bernardo [9] proposed an algorithm to derive objective priors by maximizing some entropy distances. This was further explored by Berger and Bernardo ([6], [7]) and named by them the reference priors. We obtain that the one-at-a-time reference prior for the parameter ordering $\{\rho, \sigma^2, \boldsymbol{\beta}\}$ or $\{\rho, \boldsymbol{\beta}, \sigma^2\}$ is given by

$$\pi_R(\rho, \sigma^2, \boldsymbol{\beta}) \propto (\sigma^2)^{-1}(1-\rho)^{-1}(1+(k-1)\rho)^{-1}, \tag{3.5}$$

which is exactly the same as the reference prior identified by Ghosh and Heo [37], because their model is just a special case of model in (3.1) when we set $\mathbf{X}_1 = \cdots = \mathbf{X}_n$. In addition, it can be shown that the prior in (3.5) is a second-order matching prior because it achieves approximate frequentist validity of the posterior quantiles of the interest parameter ρ with a margin of error of $o(n^{-1})$. We refer the interested readers to Datta and Ghosh ([24], [25]) and Datta and Mukerjee [26] about the second-order matching criterion in detail. The resulting marginal posterior density of ρ under this prior, denoted by $\pi_R(\rho \mid D)$, is given by

$$\pi_R(\rho \mid D) \propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1}|\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2}. \tag{3.6}$$

Given that neither π_J in (3.3) nor π_R in (3.5) is proper, it is important to study the propriety of their corresponding posterior distributions, which is summarized in the following theorem with proofs given in the Appendix B.1.

Theorem 2 *Consider the intraclass linear model in (3.1). Under either the Jeffreys prior π_J in (3.3) or the reference prior π_R in (3.5) for the unknown parameters, the joint posterior distribution of $(\rho, \sigma^2, \boldsymbol{\beta})$ is proper when $k \geq 2$.*

As commented by Bernardo [12], the problems of hypothesis testing and point estimation can be viewed as a special decision problem from a Bayesian decision-theoretic point of view. The choice of the loss function plays a central role in the statistical decision theory. There are numerous loss functions, such as the squared error loss, the zero-one loss, and the absolute error loss, whereas many of them often lack of the invariance property required in practice. For example, the squared error loss is often overused in statistical inference as a measure of the discrepancy between two sampling distributions, heavily depending on the chosen parameterizations (Bernardo [11]). In this paper, we consider the intrinsic discrepancy as a loss function due to its various appealing properties discussed in the next section.

3.3 Bayesian reference criterion

In this section, we propose an objective Bayesian solution based on the Bayesian reference criterion (BRC) proposed by Bernardo and Rueda [15]. In Subsection 3.3.1, we overview the BRC and derive the intrinsic discrepancy for the hypothesis testing of ρ. We then obtain Bayesian intrinsic statistic in Subsection 3.3.2 and Bayesian intrinsic estimator of ρ in Subsection 3.3.3.

3.3.1 Intrinsic discrepancy loss function

Without loss of generality, we assume that the probabilistic behavior of observable data $\mathbf{y}$ can be appropriately described by the probability model

$$M \equiv \Big\{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}),\ \mathbf{y} \in \mathbf{Y},\ \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \boldsymbol{\omega} \in \boldsymbol{\Omega}\Big\}, \tag{3.7}$$

where $\boldsymbol{\theta}$ is the parameter of interest and $\boldsymbol{\omega}$ is a nuisance parameter. We aim at deciding whether or not to treat the reduced model $p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega})$ under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ as a proxy for the general model M. In other words, we decide whether the model under H_0 is compatible with the observable data. Since the Kullback-Leibler (KL) direct divergence is a good measure of discrepancy between two probability distributions (Robert [62]), Bernardo [10] developed the logarithmic discrepancy derived by minimizing this divergence measure. Given that the logarithmic discrepancy is not symmetric and this feature may be unsuitable in some contexts, Bernardo and Rueda

[15] developed a symmetric version, often called the intrinsic discrepancy given by

$$\delta(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\theta}_0) = \min\Big\{\kappa(\boldsymbol{\theta}_0 \mid \boldsymbol{\omega}, \boldsymbol{\theta}),\ \kappa(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \boldsymbol{\theta}_0)\Big\},$$

where

$$\kappa(\boldsymbol{\theta}_0 \mid \boldsymbol{\omega}, \boldsymbol{\theta}) = \inf_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} \int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}) \log \frac{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega})}{p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0)}\, d\mathbf{y},$$

and

$$\kappa(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \boldsymbol{\theta}_0) = \inf_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} \int p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0) \log \frac{p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0)}{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega})}\, d\mathbf{y}.$$

The unit of the intrinsic discrepancy is the nat of information, while it could be a bit of information if the logarithm was taken in base 2 instead of base e. The intrinsic discrepancy has an invariant property under one-to-one reparametrization. For a thorough discussion of other properties, see Bernardo and Rueda [15], Bernardo and Juárez [13], Bernardo [12]. In what follows, we provide the intrinsic discrepancy between two intraclass models with its derivations given in the Appendix B.2.

Theorem 3 *The intrinsic discrepancy for testing $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$, for $\rho_0 \in \big(-(k-1)^{-1}, 1\big)$ under the intraclass model in (3.1) is given by*

$$\delta(\rho_0, \rho) = \begin{cases} \kappa(\rho_0 \mid \rho) & \text{if } \rho \in \big(-\frac{1}{k-1}, \rho_0\big] \\ \kappa(\rho \mid \rho_0) & \text{if } \rho \in (\rho_0, 1) \end{cases}, \tag{3.8}$$

where

$$\begin{aligned}\kappa(\rho \mid \rho_0) =& \frac{nk}{2}\log\left\{\frac{1+(k-2)\rho-(k-1)\rho_0\rho}{(1+(k-1)\rho)(1-\rho)}\right\} \\ &-\frac{n}{2}\log\left\{\frac{(1+(k-1)\rho_0)(1-\rho_0)^{k-1}}{(1+(k-1)\rho)(1-\rho)^{k-1}}\right\}.\end{aligned} \tag{3.9}$$

It can be easily verified that $\rho_0 \mapsto \delta(\rho_0, \rho)$ is a continuous convex function with a unique minimum at $\rho = \rho_0$. Figure 3.1 depicts the curves $\rho_0 \mapsto \delta(\rho_0, \rho)$ for $n = 1$, $k = 4$ and $\rho \in \{-0.3, 0, 0.3\}$. We observe that the corresponding curve of the intrinsic discrepancy always vanishes at $\rho_0 = \rho$.

3.3.2 Bayesian intrinsic statistic

If we select the intrinsic discrepancy as the loss function, then the intrinsic statistic can be defined as the posterior expectation of the intrinsic discrepancy loss, namely,

$$d(\rho_0 \mid D) = \int_{\Theta} \delta(\rho, \rho_0)\pi_\delta(\rho \mid D)\, d\rho, \tag{3.10}$$

where $\pi_\delta(\rho \mid D)$ is the marginal posterior distribution for ρ under the δ-reference prior when the quantity of interest is $\delta(\rho_0, \rho)$ in (3.8). Because $\delta(\rho_0, \rho)$ is a one-to-one piecewise function of ρ, we follow Proposition 1 of Bernardo [10] and show that the δ-reference prior corresponding to the parameter of interest $\delta(\rho_0, \rho)$ is exactly the same as the reference prior for ρ corresponding to the parameter of interest ρ. In addition, the posterior distribution of ρ is invariant under this kind of transformations (Bernardo and Smith [16], p. 326). The intrinsic statistic in (3.10) can thus be rewritten as

$$\begin{aligned} d(\rho_0 \mid D) &= \int_\Theta \delta(\rho_0, \rho)\pi_\delta(\rho \mid D)\, d\rho = \int_\Theta \delta(\rho_0, \rho)\pi(\rho \mid D)\, d\rho \\ &= \int_{-1/(k-1)}^{\rho_0} \kappa(\rho_0 \mid \rho)\pi(\rho \mid D)\, d\rho + \int_{\rho_0}^{1} \kappa(\rho \mid \rho_0)\pi(\rho \mid D)\, d\rho, \end{aligned}$$

where $\pi(\rho \mid D)$ is the marginal posterior distribution of ρ under either π_J in (3.3) or π_R in (3.5). We observe from Bernardo [12] that the intrinsic statistic can be interpreted as the expected value of the log-likelihood ratio against the simplified model under H_0. On the other hand, the BRC can be defined as

$$\text{Reject } H_0: \ \rho = \rho_0 \text{ when } d(\rho_0 \mid D) > d^*$$

for some given utility constant d^*. In this paper, we advocate the conventional choices $d^* \in \big\{ \log(10), \log(100), \log(1000) \big\}$ for scientific communication. The value of about $\log(10)$ indicates some evidence against H_0; the value of about $\log(100)$ provides rather strong evidence against H_0, while the value of about $\log(1000)$ can be safely used to reject H_0. For further details about these values, we refer the interested readers to Bernardo and Rueda [15], Bernardo and Juárez [13], Bernardo and Pérez [14], and Bernardo [12].

3.3.3 Bayesian intrinsic estimator

We follow Bernardo and Juárez [13] and define the intrinsic estimator of ρ as

$$\rho^* = \rho^*(D) = \arg\min_{\rho_0 \in \Theta} d(\rho_0 \mid D), \tag{3.11}$$

which is the value minimizing the posterior expectation of the intrinsic discrepancy loss function. The intrinsic estimator inherits the invariance property of the intrinsic statistic under one-to-one piecewise transformation, which means that if $\psi = \psi(\rho)$ is a one-to-one reparametrization of ρ, then the intrinsic estimator of ψ is simply $\psi^* = \psi(\rho^*)$.

3.4 Examples

We examine the performance of the proposed solution to both hypothesis testing and point estimation problems of ρ though simulation studies (Subsection 3.4.1) and a real-data application (Subsection 3.4.2).

3.4.1 Simulation study

We conduct simulation studies to investigate the behavior of the proposed solution under different scenarios. There are n observations and 2 regressors ($p = 3$) and the data are generated from the model in (3.1). Without loss of generality, we set $\sigma^2 = 1$, $\boldsymbol{\beta} = (1, 1, 1)'$ and $\boldsymbol{\Sigma} = (1 - \rho_T)\mathbf{I}_3 + \rho_T\mathbf{J}_3$, where ρ_T is the prespecified true value of ICC. Each element of $\mathbf{X}_i$ for $i = 1, \cdots, n$ is generated from a uniform density over the interval $(-2, 2)$. To check the variations of the proposed approach, ρ_T is taken to be one of four different values: -0.3, 0, 0.3, 0.8 corresponding to the correlation being negative, zero, medium, and large, respectively, while considering different sample sizes $n = 5$ (small) and $n = 20$ (medium). For each simulation setting, we consider $N = 10,000$ replications. We analyze the averaged estimates along with the mean absolute errors (MAE) given by

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|\hat{\rho}_j - \rho_T\right|,$$

where $\hat{\rho}_j$ represents the estimate of ρ_T in jth replication.

The MAEs of the Bayesian estimations and the MLE (Paul [59]) are reported in Tables 3.1 and 3.2. Several features can be drawn as follows. (i) The intrinsic estimator under π_R outperforms the one under π_J in most cases, especially when the sample size is small, and they behave similarly as n increases. (ii) The intrinsic estimator under each prior outperforms the posterior mode and is comparable with the posterior median. (iii) When the true value ρ_T is near by 0, the MLE performs the best, whereas when ρ_T is far from 0 (e.g., $\rho_T = 0.8$), the intrinsic estimator performs the best among all the estimators under consideration. (iv) On average, the MAEs of all the estimators decrease significantly with an increasing sample size. In a marked contrast with other estimators, the intrinsic one is invariant under one-to-one transformation, which is not shared others, such as the posterior mean. Simulations with other choices of ρ have also been conducted, and similar conclusions are achieved and thus not presented here for simplicity.

We further compare the frequentist coverage probability of the posterior distributions of ρ under π_J and π_R. Following Sun and Ye [64], we let α be the left tail probability and $\rho^{(\alpha)}$ be the corresponding quantile of the marginal posterior distribution $\pi(\rho \mid D)$ under either π_J or π_R. Theoretically, it follows $F(\rho^{(\alpha)}) = \int_{-\infty}^{\rho^{(\alpha)}} \pi(\rho \mid D) d\rho = \alpha$. Letting $P(\alpha \mid \rho_T) = P(\rho < \rho^{(\alpha)} \mid \rho_T, D) = P(F(\rho) < \alpha \ \mid \rho_T, D) = P(\int_{-\infty}^{\rho} \pi(\rho \mid D) d\rho < \alpha \ \mid \rho_T, D)$, we observe that $P(\alpha \mid \rho_T)$ should be very close to α if the chosen prior performs well with respect to the probability matching criterion. Table 3.3 shows the estimated tail probabilities of the posterior distributions between two priors under different scenarios. We observe that the tail probabilities of the posterior distribution of ρ under π_R are closer to the frequentist coverage probabilities than the ones under π_j. This observation is reasonable, because π_R is a second-order matching prior if ρ is the parameter of interest.

In addition to the parameter estimation, the proposed solution can be used to test any value of $\rho = \rho_0 \in (-0.5, 1)$ since $k = 3$ in our simulation study. For illustrative purposes, suppose that we are interested in evaluating whether the data are compatible with $H_0 : \rho = 0$. We analyze the frequentist behavior of the proposed solution under π_R for the hypothesis testing of ρ based on two scenarios discussed below.

First, consider the scenario in which $H_0 : \rho = 0$ is true. We simulate $5,000$ random samples from the model in (3.1) with $\rho_T = 0$ based on the simulation setup above. Figure 3.2 depicts the sampling distribution of $d(\rho \mid D)$ from the $5,000$ simulations. For $n = 5$, the significance level is around 13.24% for $d^* = \log(10)$ (mild evidence); the significance level is around 3.26% for $d^* = \log(100)$ (strong evidence), and the significance level is around 0.88% for $d^* = \log(1000)$ (safe to reject H_0). We observe that as n increases ($n = 20$), the significance level approximately goes down to 5.20%, 0.26% and 0.06%, respectively. As one would expect, the significance level significantly decreases as n increases from a frequentist viewpoint.

Second, consider the scenario in which $H_0 : \rho = 0$ is not true. We study the behavior of the sampling distribution of the proposed solution and the relative frequency of the rejection of H_0. We again simulate $5,000$ random samples from the model in (3.1) with $\rho_T \in \{-0.3, 0.3, 0.8\}$. Figure 3.3 shows the sampling distribution of $d(\rho \mid D)$ from the $5,000$ simulations. Note that the power of the proposed approach increases when ρ_T is far from the testing value $\rho_0 = 0$ or n is larger. For instance, when $H_0 : \rho = 0$ while $\rho_T = 0.8$, for $n = 5$, the relative frequency of rejecting H_0 is approximately equal to 79.46% for $d^* = \log(10)$, to 35.32% for $d^* = \log(100)$, and to 6.56% for $d^* = \log(1000)$; for $n = 20$, this relative frequency significantly increases to 100%, 99.84%, and 98.68%, respectively. We may thus conclude that the power of the proposed solution increases with n and that the performance of the proposed solution is quite satisfactory for the problems of hypothesis testing and point estimation of ρ in the intraclass model in (3.1).

Given that there are two objective priors: the reference prior (π_R) or the Jeffreys prior (π_J), which of them is preferable for the proposed solution in practical applications? Numerical evidence from the above simulation studies showed that the Bayesian estimations under π_R outperform the ones under π_R. Additionally, π_R is also a second-order matching prior if ρ is the parameter of interest. We thus have a preference to recommend the use of π_R in the analysis of the ICC.

3.4.2 An illustrative example

We use a real data example to illustrate the practical application of the proposed solution. The orthodontic data set is present in Table 3.4 and obtained from Chapter

5.2 of Frees [33]: twenty-seven individuals including 16 boys and 11 girls were measured for distances from the pituitary to the pteryomaxillary fissure in millimeters, at ages 8, 10, 12, and 14. We consider the intraclass model of the form

$$\mathbf{y}_i = \beta_0\mathbf{j}_4 + \beta_1\mathbf{A}_i + \beta_2\mathrm{G}_i\mathbf{j}_4 + \boldsymbol{\varepsilon}_i, \quad i = 1, \cdots, 27,$$

where $\mathbf{y}_i = (y_{i1},\ y_{i2},\ y_{i3},\ y_{i4})^T$ with y_{ij} being the distance for individual i measured at age j, $\mathbf{A}_i = (8,\ 10,\ 12,\ 14)^T$ is a 4×1 vector of ages and G_i represents the gender (1 for male and 0 for female), and $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_4, \sigma^2\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_4 + \rho\mathbf{J}_4$. We observe from Figure 3.4(a) that the marginal posterior densities for ρ under two objective priors are quite normal in shape. Table 3.5 provides the point estimators for ρ under different procedures. We here analyze the results under π_R for simplicity. The intrinsic estimator $\rho^* = 0.622$ is close to the posterior median equal to 0.620, whereas both are slightly different from the MLE equal to 0.597. According to the non-rejection regions with $d^* \in \{\log(10), \log(100), \log(1000)\}$ presented in Figure 3.4(b), we somehow doubt that the true value of ρ is outside $R_{\log(10)} = (0.423,\ 0.773)$; we seriously doubt that ρ is outside $R_{\log(100)} = (0.304,\ 0.833)$, and we are almost sure that the true correlation value ρ is not outside $R_{\log(1000)} = (0.211,\ 0.870)$.

On the other hand, the proposed solution can be used for the hypothesis testing of $\rho = \rho_0 \in (-1/3, 1)$. If we are interested in testing $H_0 : \rho = \rho_0 = 0$ versus $H_1 : \rho \neq \rho_0$.

we can numerically verify that the intrinsic statistic under π_R is

$$d(\rho_0 \mid D) = \int_{-1/3}^{1} \delta(\rho_0, \rho)\pi(\rho \mid D)\, d\rho \approx 14.2747 \approx \log(1582791),$$

which indicates that the expected value of the average of the log likelihood ratio against H_0 is about 14.2747, showing that the likelihood ratio is expected to be about $1,582,791$. Thus, we may conclude that the data provide very strong evidence against H_0 and that the null hypothesis is opposed to the observable data. Due to the invariance property of the proposed solution, if the parameter of interest is ρ^3, then its intrinsic estimator is simply $(\rho^*)^3 \approx 0.622^3$, and the corresponding non-rejection regions are simply given by $\tilde{R}_{\log(10)} = (0.076,\ 0.462)$, $\tilde{R}_{\log(100)} = (0.028,\ 0.578)$, and $\tilde{R}_{\log(1000)} = (0.009,\ 0.659)$, respectively.

3.5 Concluding remarks

In this paper, we first derived two objective priors for the unknown parameters in the intraclass model in (3.1) and proved that both result in proper posterior distributions. Within a Bayesian decision-theoretic framework, we then proposed an objective Bayesian solution to both hypothesis testing and point estimation problems of the ICC ρ. The proposed solution has an appealing invariance property under one-to-one reparametrization of the quantity of interest, which is not shared by some commonly used estimators, such us the posterior mean.

It deserves mentioning that the proposed solution can be directly applied to the balanced one-way random effect ANOVA model, since it is a special case of the intraclass model in (3.1) if we let $\sigma^2 = \sigma_a^2 + \sigma_e^2$ and $\rho = \sigma_a^2/\sigma^2 \in (0, 1)$, where σ_a^2 and σ_e^2 stand for the treatment and error variances, respectively. This observation motivates a possible extension of the proposed solution to the unbalanced model with different number of observations in each class, which is currently under investigation and will be reported elsewhere.

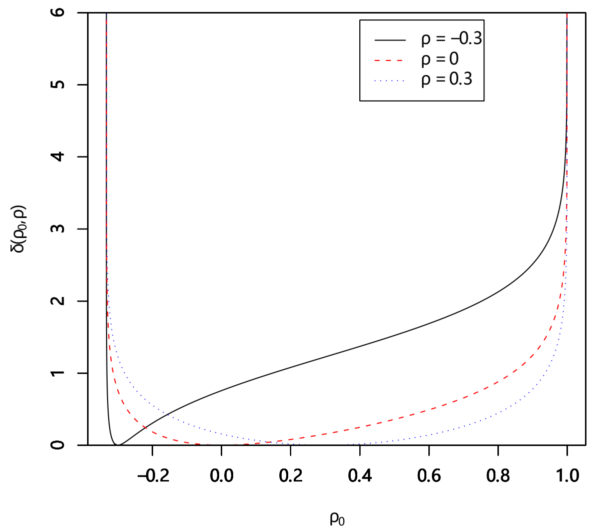


Figure 3.1: The intrinsic discrepancy $\delta(\rho_0, \rho)$ in (3.8) as a function of ρ_0 for $n = 1$, $k = 4$ and $\rho \in \{-0.3, 0, 0.3\}$

Table 3.1

The MAE of the Bayesian estimators for ρ based on 10,000 replications in the simulation study.

		n=5				n=20			
ρ_T	Prior	Intrinsic	Mean	Median	Mode	Intrinsic	Mean	Median	Mode
-0.3	π_R	0.148	0.155	0.149	0.162	0.058	0.060	0.058	0.059
	π_J	0.164	0.163	0.166	0.185	0.060	0.060	0.060	0.062
0	π_R	0.242	0.213	0.243	0.333	0.108	0.105	0.108	0.115
	π_J	0.294	0.263	0.296	0.379	0.114	0.111	0.114	0.121
0.3	π_R	0.268	0.230	0.268	0.377	0.119	0.115	0.119	0.129
	π_J	0.315	0.276	0.315	0.412	0.124	0.119	0.124	0.133
0.8	π_R	0.148	0.157	0.148	0.151	0.057	0.059	0.057	0.057
	π_J	0.142	0.141	0.141	0.153	0.056	0.056	0.056	0.058

Table 3.2
The MAE of the MLE for ρ based on $10,000$ replications in the simulation study.

	n=5	n=20
-0.3	0.137	0.058
0	0.198	0.102
0.3	0.231	0.118
0.8	0.236	0.074

Table 3.3
The estimated tail probabilities of posterior distributions based on $10,000$ replications in the simulation study.

		n=5		n=20	
ρ_T	Prior	$P(0.05 \mid \rho_T)$	$P(0.90 \mid \rho_T)$	$P(0.05 \mid \rho_T)$	$P(0.90 \mid \rho_T)$
-0.3	π_R	0.0453	0.9127	0.0477	0.9010
	π_J	0.0497	0.9145	0.0425	0.9166
0	π_R	0.0460	0.9069	0.0535	0.8977
	π_J	0.0842	0.8598	0.0617	0.8879
0.3	π_R	0.0439	0.9119	0.0453	0.9054
	π_J	0.1021	0.8357	0.0614	0.8816
0.8	π_R	0.0441	0.9087	0.0484	0.9001
	π_J	0.1341	0.7825	0.0779	0.8587

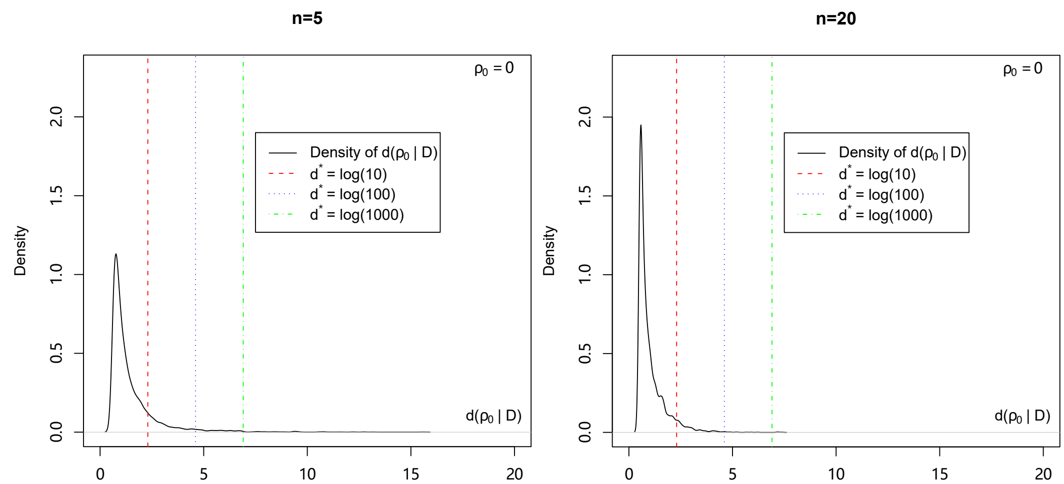


Figure 3.2: Sampling distribution of $d(\rho \mid D)$ under H_0 obtained from the 5,000 simulations with $\rho_T = 0$ for different sample sizes when testing $H_0 : \rho = 0$.

Table 3.4
The orthodontic data from Frees [33].

	Age of girls				Age of boys			
Number	8	10	12	14	8	10	12	14
1	21	20	21.5	23	26	25	29	31
2	21	21.5	24	25.5	21.5	22.5	23	26.5
3	20.5	24	24.5	26	23	22.5	24	27.5
4	23.5	24.5	25	26.5	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	20	23.5	22.5	26
6	20	21	21	22.5	24.5	25.5	27	28.5
7	21.5	22.5	23	25	22	22	24.5	26.5
8	23	23	23.5	24	24	21.5	24.5	25.5
9	20	21	22	21.5	23	20.5	31	26
10	16.5	19	19	19.5	27.5	28	31	31.5
11	24.5	25	28	28	23	23	23.5	25
12					21.5	23.5	24	28
13					17	24.5	26	29.5
14					22.5	25.5	25.5	26
15					23	24.5	26	30
16					22	21.5	23.5	25

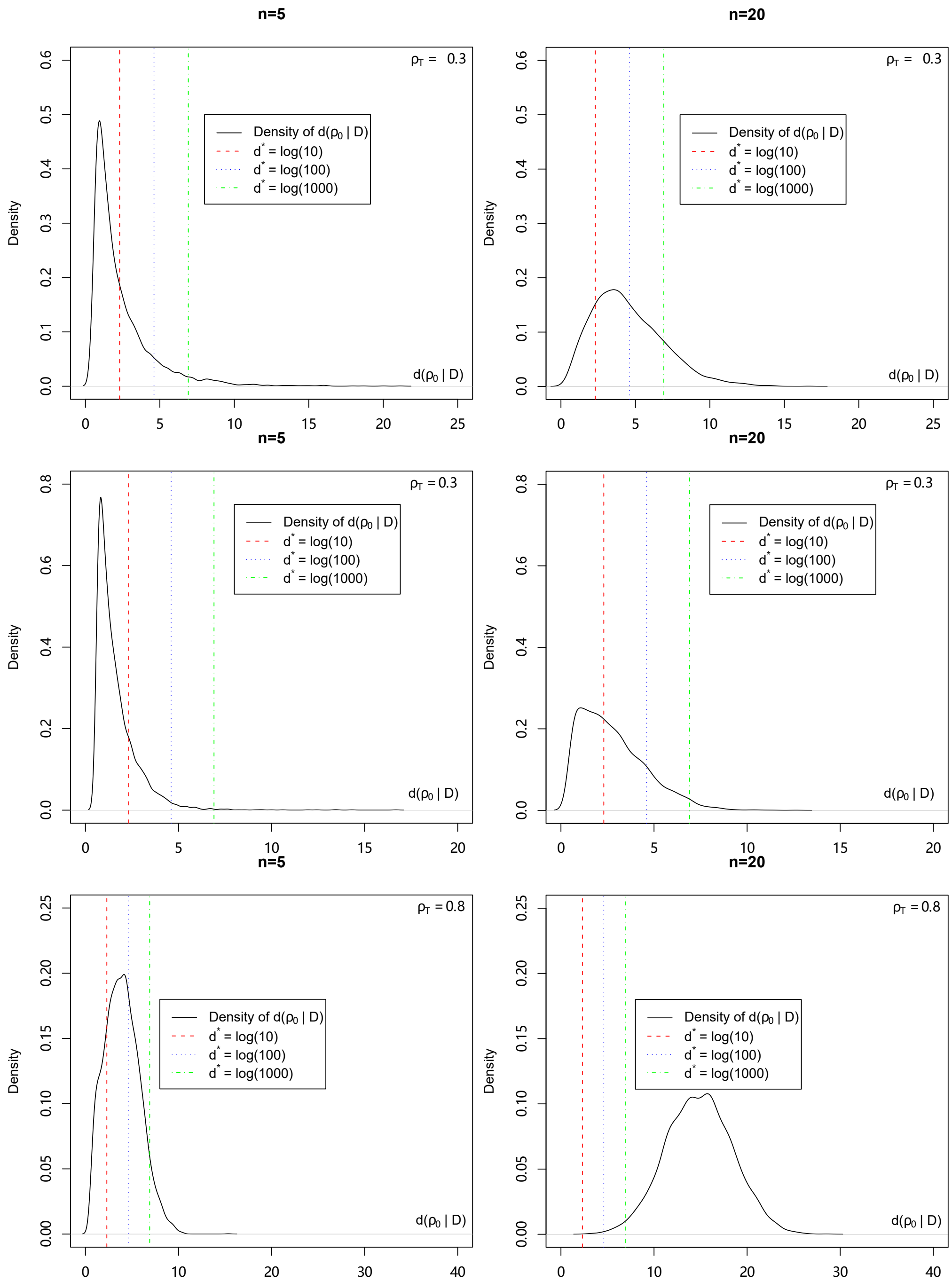
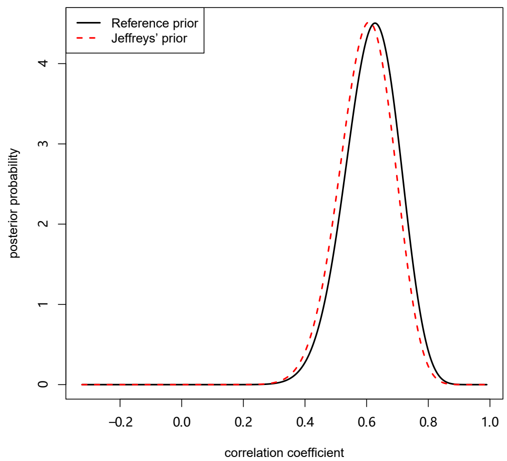
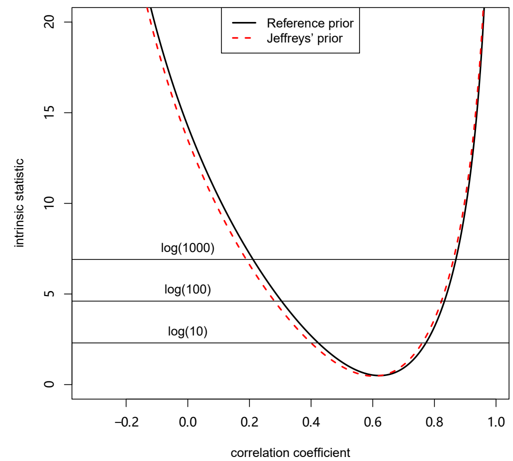


Figure 3.3: Sampling distribution of $d(\rho \mid D)$ under H_0 obtained from 5,000 simulations with $\rho_T \in \{-0.3, 0.3, 0.8\}$ for different sample sizes when testing $H_0 : \rho = 0$.



(a) Marginal posterior distribution



(b) Intrinsic statistic

Figure 3.4: The marginal posterior density for ρ based on two objective priors (left), and the intrinsic statistic with the non-rejection regions corresponding to the threshold values $d^* \in \{\log(10), \log(100), \log(1000)\}$ (right) for the orthodontic data in Frees [33].

Table 3.5
Estimations of ρ for the orthodontic data from Frees [33].

Priors	Intrinsic	Mean	Median	Mode
π_J	0.603	0.598	0.601	0.608
π_R	0.622	0.616	0.620	0.627

Chapter 4

Bayesian Quantile Regression for Semiparametric Mixed-Effects Models

Semiparametric mixed-effects models (SPMMs) are widely used for longitudinal data in various practical applications. The model contains a linear part modeling some explanatory variables and a nonlinear part associated with a time effect. As quantile regression has become a popular tool in data analysis, in this article, we aim to put forward Bayesian quantile regression for SPMMs. The quantile structure is attained by specifying the error term as the asymmetric Laplace distribution (ALD). A cubic spline approximation is applied for the nonlinear part. We model the variation within subjects by specifying the variance of random effect as a function of some explanatory variables. An efficient Gibbs sampler with the Metropolis-Hastings algorithm is developed to sample the parameters from their posterior distributions. Simulation studies and a real data example are used to illustrate the proposed methodology.

4.1 Introduction

Quantile regression is a type of regression analysis and has been widely applied in a wide range of disciplines, such as biological studies, econometrics, fiance, and social sciences. Quantile regression provides information that the classical mean regression cannot reflect, because it quantifies the association of explanatory variables with a specific quantile of a dependent variable and is also quite insensitive to heteroscedasticity and outliers, which often occur in many practical applications. Since the seminal work of Koenker and Bassett [48], quantile regression has drawn increasingly attention in the literature from both Bayesian and frequentist points of view; see, for example, Kotz et al. [52], Yu and Moyeed [70], Alhamzawi and Ali [2], to name just a few.

It deserves mentioning that due to a close link between the asymmetric Laplace distribution (ALD) and quantile regression, the quantile regression analysis has attracted a great deal of attention from a Bayesian perspective. Of particular note is that Yu and Moyeed [70] showed that the problem of estimating quantile regression coefficients in the linear quantile regression model is equivalent to the one of maximizing the likelihood function in terms of the regression coefficients by specifying the ALD as the error distribution. Thereafter, numerous researchers conducted Bayesian quantile regression analysis with the use of the ALD as the error distribution.

Kottas and Krnjaji [51] proposed a Bayesian semiparametric methodology for quantile regression using Dirichlet process mixtures for the error distribution. Kozumi and Kobayashi [53] developed an efficient Gibbs sampling algorithm for Bayesian quantile regression based on a location-scale mixture representation of the ALD. Chen and Yu [20] studied Bayesian inference for nonparametric quantile regression and adopted piecewise polynomial functions for curve fitting. Wang [65] proposed nonlinear mixed-effects models based on a likelihood-based approach using the ALD. We observe that the ALD provides a natural and effective way in Bayesian quantile regression framework. The estimators can be sampled from their posterior distributions through highly efficient Markov chain Monte Carlo (MCMC) algorithms.

Bayesian quantile regression of mean models has been extensively studied in the literature, whereas the problem of jointly modeling mean and variance has often been relatively neglected. Modeling variance is often necessary to obtain confidence intervals and other predictions. Aitkin [1] proposed maximum likelihood (ML) estimation for mean and variance in normal liner regression models. Cepeda and Gamerman [19] studied variance heterogeneity of normal regression models from a Bayesian perspective. In Lombardia and Sperlich [58], the ML estimation is applied to the generalized mixed-effects model and the variance of random effects is estimated to improve statistical inference on estimating parameters. Since semiparametric mixed-effects models are useful tools for analyzing longitudinal data, Alhamzawi and Ali [2] derived Bayesian quantile regression for longitudinal data and Xu et al. [67] presented a Bayesian approach for semiparametric mixed-effects models. In this paper, we generalize these methods into Bayesian quantile regression on semiparametric mixed-effects models. In particular, the variance of random effects is modeled as a function of the explanatory variables.

The remainder of this chapter is organized as follows. In section 4.2, we present the model in matrix form. In section 4.3, we specify prior distributions of the parameters and derive their corresponding posteriors. The Gibbs sampling algorithm is implemented to obtain the estimated parameters. Simulation study and a real data example are illustrated in section 4.4 and section 4.5, which evaluate the proposed methodology and apply it to practical use. Some discussion are given in section 4.6.

4.2 Bayesian quantile structure for semiparametric mixed-effects models

In this section, we discuss semiparametric mixed-effects double regression models and adopt the B-spline method to fit semeiparametric models in Subsection 4.2.1. We then consider Bayesian quantile analysis of semiparametric mixed-effects double regression models in Subsection 4.2.2.

4.2.1 Semiparametric mixed-effects models

Consider the semiparametric mixed-effects model of the form

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + g(t_{ij}) + v_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, m_i, \tag{4.1}$$

where y_{ij} is the response variable of the ith subject on the jth measurement, $\mathbf{x}_{ij} = (x_{ij1}, \cdots, x_{ijp})^T$ is a $p \times 1$ vector of predictor variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown common regression coefficients, $g(t_{ij})$ is an unknown smooth function associated with a univariate observed covariate t_{ij} associated with time, v_i is a random effect of each subject with $v_i \sim \mathrm{N}(0, \sigma_i^2)$ with σ_i^2 being the heterogeneity variance of the random effect, and ε_{ij} is the error term. Here, the superscript T represents the transpose of a matrix or a vector. For practical applications, we often assume that there exists a variance heterogeneity of each subject and that the variance σ_i^2 is related to several predictors $\mathbf{z}_i = (z_{i1}, \cdots, z_{iq})^T$ such that $\sigma_i^2 = h(\mathbf{z}_i, \boldsymbol{\gamma})$, where $h(\cdot, \cdot)$ is a known function to model varying variance and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_q)^T$ is a $q \times 1$ vector of regression coefficients. There are several known forms for $h(\mathbf{z}_i, \boldsymbol{\gamma})$, such as log-linear model or power product model; see Xu et al. [67].

To explicitly specify the model in (4.1), we adopt the B-spline technique to approximate the nonparametric function $g(\cdot)$, which converts the nonparametric function into a linear function consisting of a set of basis functions. Without loss of generality, we assume that $t_{ij} \in [0, 1]$, which can be partitioned as $0 = s_0 < s_1 < \cdots < s_{k_n} < s_{k_n+1} = 1$, where $\{s_i\}$ is an internal knot. This provides that there are $K = k_n + M$ normalized B-spline basis functions $\{\pi_k(t_{ij})\}$ of order M that form a basis the linear spline space, where $\pi_k(\cdot)$ is the k-th basis function and $k = 1, 2, \cdots, K$. In this paper, we consider the cubic splines (i.e., $M = 4$), because they have two continuous derivatives which are often sufficient to give smooth approximations and a third degree piecewise polynomial usually behaves numerically well. In addition, by following He and Fung [41], we choose the number of knots to be the integer part of $N^{1/5}$, where $N = \sum_{i=1}^n m_i$. Thus, the model (4.1) can be linearized as

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{b}_{ij}^T \boldsymbol{\alpha} + v_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, m_i, \tag{4.2}$$

where $\mathbf{b}_{ij} = (\pi_1(t_{ij}), \cdots, \pi_K(t_{ij}))^T$ is a $K \times 1$ vector of basis functions and $\boldsymbol{\alpha}$ is a $K \times 1$ vector of the regression coefficients for the basis functions.

According to Kozumi and Kobayashi [53], we can obtain the τ-th quantile regression estimator for $\boldsymbol{\beta}_\tau$ and $\boldsymbol{\alpha}_\tau$ by minimizing the following objective loss function

$$(\hat{\boldsymbol{\beta}}_\tau,\ \hat{\boldsymbol{\alpha}}_\tau) = \arg \min_{\boldsymbol{\beta}_\tau, \boldsymbol{\alpha}_\tau} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \rho_\tau \big(y_{ij} \ - \ \mathbf{x}_{ij}^T \boldsymbol{\beta}_\tau \ - \ \mathbf{b}_{ij}^T \boldsymbol{\alpha}_\tau \ - \ v_i \big), \tag{4.3}$$

where $\tau \in (0, 1)$ is a given quantile level and $\rho_\tau(\cdot)$ is the check loss function defined as

$$\rho_\tau(u) = u\{\tau - I(u < 0)\},$$

where $I(\cdot)$ denotes the indicator function. Given that the estimators cannot be obtained by differentiating the objective function in (4.3), we may interior point methods to calculate these quantile regression estimators; see, for example, Koenker and Park [50]. We observe from Yu and Moyeed [70] that the minimization problem in (4.3) is equivalent to maximizing the likelihood function by specifying the error term ε_{ij} in (4.2) as the asymmetric Laplace distribution (ALD). This relationship has been widely adopted to develop Bayesian quantile regression methods in the literature; see, for example, Kozumi and Kobayashi [53].

4.2.2 Bayesian quantile regression models

To fully conduct Bayesian quantile analysis for the model in (4.2), we may assume that the error term ε_{ij} follows the ALD, which can be written as a scale mixture of normals with the scale mixing parameter following an exponential distribution summarized in the following proposition.

Proposition 2 *Let* $e \sim \mathrm{Exp}(\theta^{-1})$ *and* $r \sim \mathrm{N}(0,1)$ *be two independent random variables. Then* $\varepsilon \sim \mathrm{ALD}(\mu,\theta,\tau)$ *can be presented by*

$$\varepsilon = \mu + k_1 e + r\sqrt{k_2 \theta e},$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ *and* $k_2 = \frac{2}{\tau(1-\tau)}$.

Based on this mixture representation in Proposition 2, the model (4.2) with the random error term $\varepsilon_{ij} \sim \mathrm{ALD}(0,\theta,\tau)$ can be rewritten as

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{b}_{ij}^T \boldsymbol{\alpha} + v_i + k_1 e_{ij} + r_{ij}\sqrt{k_2 \theta e_{ij}}, \tag{4.4}$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $k_2 = \frac{2}{\tau(1-\tau)}$, e_{ij} and r_{ij} are random variables independent of each other, such that $e_{ij} \sim \mathrm{Exp}(\theta^{-1})$ and $r_{ij} \sim \mathrm{N}(0,1)$.

For notational simplicity, let $\mathbf{y} = (\mathbf{y}_1^T, \cdots, \mathbf{y}_n^T)^T$ be the vector of all response observations with $\mathbf{y}_i = (y_{i1}, \cdots, y_{im_i})^T$, $\mathbf{t} = (\mathbf{t}_1^T, \cdots, \mathbf{t}_n^T)^T$ be the time sequence vector with $\mathbf{t}_i = (t_{i1}, \cdots, t_{im_i})^T$, $\mathbf{X} = (X_1^T, \cdots, X_n^T)^T$ be the design matrix with $X_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{im_i})^T$, $\mathbf{B} = (\mathbf{B}_1^T, \cdots, \mathbf{B}_n^T)^T$ with $\mathbf{B}_i = (\mathbf{b}_{i1}, \cdots, \mathbf{b}_{im_i})^T$, $\mathbf{e} = (\mathbf{e}_1^T, \cdots, \mathbf{e}_n^T)^T$ with $\mathbf{e}_i = (e_{i1}, \cdots, e_{im_i})^T$, $\mathbf{r} = (\mathbf{r}_1^T, \cdots, \mathbf{r}_n^T)^T$ with $\mathbf{r}_i = (r_{i1}, \cdots, r_{im_i})^T$. Denote $\tilde{\mathbf{v}} = (\mathbf{v}_1^T, \cdots, \mathbf{v}_n^T)^T$ with $\mathbf{v}_i = v_i \otimes \mathbf{1}_{m_i}$, where $\otimes$ is the Kronecker product and $\mathbf{1}_{m_i}$ is a vector which has m_i 1s. Then the model (4.4) can be written as a matrix form given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\alpha} + \tilde{\mathbf{v}} + k_1\mathbf{e} + \mathbf{r} \circ \sqrt{k_2\theta\mathbf{e}}, \tag{4.5}$$

where $\circ$ is the Hadamard product, which lets two vectors of the same dimensions multiply element by element.

The likelihood function of all model parameters is given by

$$\begin{aligned} &L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \mathbf{v}, \mathbf{e} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}} |\mathbf{E}|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{E}^{-1} (\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2}\mathbf{v}^T \boldsymbol{\Sigma}^{-1} \mathbf{v} \Big\}, \end{aligned} \tag{4.6}$$

where $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)^T$ with $\mathbf{z}_i = (z_{i1}, \cdots, z_{iq})^T$, $\mathbf{E} = k_2\theta\mathrm{diag}(\mathbf{e}^T)$, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\alpha} + \tilde{\mathbf{v}} + k_1\mathbf{e}$, and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_n^2)$. The representation of the likelihood function in (4.6) allows us to develop an easy way to construct an efficient Gibbs sampler algorithm for the posterior sampling in the following section.

4.3 Posterior inference

Bayesian analysis begins with the prior specifications for the unknown model parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)$ in (4.6). For the unknown parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, we assume that they are independently distributed as multivariate normal distributions such that $\boldsymbol{\alpha} \mid \phi^2 \sim \mathrm{N_K}(\boldsymbol{\alpha}_0, \phi^2 \mathbf{I}_K)$, $\boldsymbol{\beta} \mid \theta \sim \mathrm{N_p}(\boldsymbol{\beta}_0, \theta \mathbf{B}_\beta)$, and $\boldsymbol{\gamma} \sim \mathrm{N_q}(\boldsymbol{\gamma}_0, \mathbf{B}_\gamma)$, respectively, where $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$, $\mathbf{B}_\beta$, $\mathbf{B}_\gamma$ are the prespecified hyperparameters, and ϕ^2 follows an inverse Gamma distribution, denoted by $\phi^2 \sim \text{Inv-Gamma}(a_{\phi^2}, b_{\phi^2})$, with a_{ϕ^2} and b_{ϕ^2} being known positive constants. For the unknown parameter θ, we assume that it follows an inverse Gamma distribution denoted by $\theta \sim \text{Inv-Gamma}(a_\theta, b_\theta)$, where a_θ, and b_θ are known positive constants.

The joint posterior distribution, in a combination of the likelihood function in (4.6), the distribution of the latent variable $\mathbf{e}$, and the proposed prior for $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)$ is given by

$$
\begin{aligned}
& p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \mathbf{v}, \mathbf{e} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\
& \quad \propto L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, \mathbf{v}, \mathbf{e} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(\mathbf{e} \mid \theta) p(\boldsymbol{\alpha} \mid \phi^2) p(\boldsymbol{\beta} \mid \theta) p(\boldsymbol{\gamma}) p(\phi^2) p(\theta),
\end{aligned}
$$

which is not recognizable, and thus, it may be prohibitive to directly adopt numerical techniques to draw Bayesian inference for the unknown parameters. In what follows, we first obtain the full conditional distribution of each unknown parameter and then construct an efficient Gibbs sampler with the Metropolis-Hastings algorithm for the posterior samplings.

The posterior distributions of each parameter are as follows:

* For θ, since $e_{ij} \mid \theta \sim \exp(\theta^{-1})$, $\boldsymbol{\beta} \mid \theta \sim \mathrm{N_p}(\boldsymbol{\beta}_0, \theta\mathbf{B}_\beta)$, and $\theta \sim$ Inv-Gamma(a_θ, b_θ), it follows

$$\begin{aligned} & p(\theta \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \quad \propto L(\theta \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(\mathbf{e} \mid \theta) p(\boldsymbol{\beta} \mid \theta) p(\theta) \\ & \quad \propto \theta^{-a_\theta^\star - 1} \exp\left(-\frac{b_\theta^\star}{\theta}\right), \end{aligned} \tag{4.7}$$

where $a_\theta^\star = \frac{3N+p}{2} + a_\theta$ and $b_\theta^\star = \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{E}_0^{-1}(\mathbf{y} - \boldsymbol{\mu}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{B}_\beta^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{e}^T \mathbf{1}_N + b_\theta$ with $\mathbf{E}_0 = k_2 \mathrm{diag}\{\mathbf{e}^T\}$.

* For ϕ^2, since $\boldsymbol{\alpha} \mid \phi^2 \sim \mathrm{N_K}(\boldsymbol{\alpha}_0, \phi^2 \mathbf{I}_K)$, and $\phi^2 \sim$ Inv-Gamma(a_{ϕ^2}, b_{ϕ^2}), it follows

$$p(\phi^2 \mid \boldsymbol{\alpha}) \propto p(\boldsymbol{\alpha} \mid \phi^2) p(\phi^2) \propto (\phi^2)^{-a_{\phi^2}^\star - 1} \exp\left(-\frac{b_{\phi^2}^\star}{\phi^2}\right), \tag{4.8}$$

where $a_{\phi^2}^\star = \frac{K}{2} + a_{\phi^2}$ and $b_{\phi^2}^\star = \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathrm{b}_{\phi^2}$.

* For $\boldsymbol{\alpha}$, since $\boldsymbol{\alpha} \mid \phi^2 \sim \mathrm{N_K}(\boldsymbol{\alpha}_0, \phi^2\mathbf{I}_K)$, it follows

$$\begin{aligned} p(\boldsymbol{\alpha} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) &\propto L(\boldsymbol{\alpha} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(\boldsymbol{\alpha} \mid \phi^2) \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0^\star)^T \mathbf{B}_\alpha^{\star -1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0^\star)\right\}, \end{aligned} \tag{4.9}$$

where $\boldsymbol{\alpha}_0^\star = \mathbf{B}_\alpha^\star((\phi^2)^{-1}\boldsymbol{\alpha}_0 + \mathbf{B}^T\mathbf{E}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \tilde{\mathbf{v}} - k_1\mathbf{e}))$ and $\mathbf{B}_\alpha^\star = (\mathbf{B}^T\mathbf{E}^{-1}\mathbf{B} + (\phi^2)^{-1}\mathbf{I}_K)^{-1}$.

* For $\boldsymbol{\beta}$, since $\boldsymbol{\beta} \mid \theta \sim \mathrm{N_p}(\boldsymbol{\beta}_0, \theta\mathbf{B}_\beta)$, it follows

$$\begin{aligned} p(\boldsymbol{\beta} \mid \theta, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) &\propto L(\boldsymbol{\beta} \mid \theta, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(\boldsymbol{\beta} \mid \theta) \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0^\star)^T \mathbf{B}_\beta^{\star -1} (\boldsymbol{\beta} - \boldsymbol{\beta}_0^\star)\right\}, \end{aligned} \tag{4.10}$$

where $\boldsymbol{\beta}_0^\star = \mathbf{B}_\beta^\star(\theta^{-1}\mathbf{B}_\beta^{-1}\boldsymbol{\beta}_0 + \mathbf{X}^T\mathbf{E}^{-1}(\mathbf{y} - \tilde{\mathbf{v}} - \mathbf{B}\boldsymbol{\alpha} - k_1\mathbf{e}))$ and $\mathbf{B}_\beta^\star = (\mathbf{X}^T\mathbf{E}^{-1}\mathbf{X} + \theta^{-1}\mathbf{B}_\beta^{-1})^{-1}$.

* For $\mathbf{v}$, the conditional posterior is given by

$$\begin{aligned} &p(\mathbf{v} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ &\quad \propto \exp\left\{-\frac{1}{2}(\tilde{\mathbf{v}}^T\mathbf{E}^{-1}\tilde{\mathbf{v}} - 2(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha} - k_1\mathbf{e})^T\mathbf{E}^{-1}\tilde{\mathbf{v}} + \mathbf{v}^T\boldsymbol{\Sigma}^{-1}\mathbf{v})\right\} \\ &\quad \propto \exp\left\{-\frac{1}{2}(\mathbf{v} - \mathbf{v}_0^\star)^T \mathbf{B}_v^{\star -1} (\mathbf{v} - \mathbf{v}_0^\star)\right\}, \end{aligned} \tag{4.11}$$

where $\mathbf{v}_0^\star = \mathbf{B}_v^\star \mathbf{w}$, $\mathbf{B}_v^\star = (\mathbf{A} + \mathbf{\Sigma}^{-1})^{-1}$ with

$$\mathbf{A} = (k_2\theta)^{-1}\text{diag}\{\sum_{j=1}^{m_1} e_{1j}^{-1}, \cdots, \sum_{j=1}^{m_n} e_{nj}^{-1}\}$$

and $\mathbf{w} = (w_1, \cdots, w_n)^T$ with $w_i = (k_2\theta)^{-1} \sum_{j=1}^{m_j} \left\{(y_{ij} - \mathbf{x}_{ij}^T\boldsymbol{\beta} - \mathbf{b}_{ij}^T\boldsymbol{\alpha} - k_1 e_{ij})e_{ij}^{-1}\right\}$.

* For $\mathbf{e}$, since $e_{ij} \mid \theta \sim \text{Exp}(\theta)$, it follows

$$\begin{aligned} p(e_{ij} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) &\propto L(e_{ij} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{v}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(e_{ij} \mid \theta) \\ &\propto e_{ij}^{-\frac{1}{2}} \exp\left\{-(a_e^\star e_{ij} + \frac{b_{eij}^\star}{e_{ij}})/2\right\}, \end{aligned} \tag{4.12}$$

where $a_e^\star = \frac{k_1^2 + 2k_2}{k_2\theta}$ and $b_{eij}^\star = \frac{(y_{ij} - \mathbf{x}_{ij}^T\boldsymbol{\beta} - \mathbf{b}_{ij}^T\boldsymbol{\alpha} - v_i)^2}{k_2\theta}$.

* For $\boldsymbol{\gamma}$, since $\boldsymbol{\gamma} \sim \text{N}_\text{q}(\boldsymbol{\gamma}_0, \mathbf{B}_\gamma)$, it follows

$$\begin{aligned} & p(\boldsymbol{\gamma} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto L(\boldsymbol{\gamma} \mid \theta, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{v}, \mathbf{e}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) p(\boldsymbol{\gamma}) \\ & \propto |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v} - \frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{B}_\gamma^{-1}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\right\}. \end{aligned} \tag{4.13}$$

According to these conditional posteriors from (4.7) to (4.13), we construct an efficient Gibbs sampler with the Metropolis-Hastings algorithm for the posterior simulation summarized as follows.

Step 1: Set up initial values $\mathbf{\Phi}^{(0)} = (\theta^{(0)}, \phi^{2^{(0)}}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}^{(0)})$, $\mathbf{e}^{(0)}$, and $\mathbf{v}^{(0)}$.

Step 2: For the kth iteration, based on $\mathbf{\Phi}^{(k)} = (\theta^{(k)}, \phi^{2(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$, $\mathbf{e}^{(k)}$, and $\mathbf{v}^{(k)}$, update $\tilde{\mathbf{v}}^{(k)} = ((\mathbf{v}_1^{(k)})^T, \cdots, (\mathbf{v}_n^{(k)})^T)^T$ with $\mathbf{v}_i^{(k)} = v_i^{(k)} \otimes \mathbf{1}_{m_i}$; $\mathbf{\Sigma}^{(k)} = \mathrm{diag}((\sigma_1^2)^{(k)}, \cdots, (\sigma_n^2)^{(k)})^T$ and $\mathbf{E}^{(k)} = k_2 \theta^{(k)} \mathrm{diag}(\mathbf{e}^{(k)})$.

Step 3: Based on $\mathbf{\Phi}^{(k)} = (\theta^{(k)}, \phi^{2(k)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})$, $\mathbf{e}^{(k)}$, and $\mathbf{v}^{(k)}$, sample $\mathbf{\Phi}^{(k+1)} = (\theta^{(k+1)}, \phi^{2(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k+1)})$, $\mathbf{e}^{(k+1)}$, and $\mathbf{v}^{(k+1)}$ as follows:

(i) Sampling $\theta^{(k+1)} \mid \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)}, \mathbf{v}^{(k)}, \mathbf{e}^{(k)}, \mathbf{E}^{(k)}$ from Inv-Gamma$(a_\theta^\star, b_\theta^\star)$,

(ii) Sampling $\phi^{2(k+1)} \mid \boldsymbol{\alpha}^{(k)}$ from Inv-Gamma$(a_{\phi^2}^\star, b_{\phi^2}^\star)$,

(iii) Sampling $\boldsymbol{\alpha}^{(k+1)} \mid \theta^{(k+1)}, \phi^{2(k+1)}, \boldsymbol{\beta}^{(k)}, \mathbf{v}^{(k)}, \mathbf{e}^{(k)}, \mathbf{E}^{(k)}$ from $\mathrm{N_K}(\boldsymbol{\alpha}_0^\star, \mathbf{B}_\alpha^\star)$,

(iv) Sampling $\boldsymbol{\beta}^{(k+1)} \mid \theta^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \mathbf{v}^{(k)}, \mathbf{e}^{(k)}, E^{(k)}$ from $\mathrm{N_p}(\boldsymbol{\beta}_0^\star, \mathbf{B}_\beta^\star)$,

(v) Sampling $\mathbf{v}^{(k+1)} \mid \theta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \mathbf{e}^{(k)}, \boldsymbol{\gamma}^{(k)}, \mathbf{\Sigma}^{(k)}$ from $\mathrm{N_n}(\mathbf{v}_0^\star, \mathbf{B}_v^\star)$,

(vi) Sampling $\mathbf{e}_{ij}^{(k+1)} \mid \theta^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \boldsymbol{\alpha}^{(k+1)}, \mathbf{v}^{(k+1)}$ from GIG$(\frac{1}{2}, a_e^\star, b_{eij}^\star)$, where GIG represents the generalized inverse Gaussian distribution.

(vii) Sampling $\boldsymbol{\gamma}^{(k+1)} \mid \mathbf{v}^{(k+1)}$ from (4.13) using an efficient Metropolis-Hastings method in the Appendix C.

Step 4: Repeating Steps 2 and 3 until the specified number of iterations, i.e., $k = J$.

The posterior sampling algorithm above was conducted in R language and can be made available upon request to the corresponding author.

4.4 Simulation study

In this section, we conduct some simulation studies to evaluate the finite sample performances of the proposed Bayesian quantile semiparametric approach with respect to the different choices of prior information (Subsection 4.4.1) and different non-standard error distributions (Subsection 4.4.2). All the simulation results were based on 10,000 iterations with discarding the first 2000 as the burn-in period. There is no evidence of lack of convergence in MCMC simulation according to the run length control diagnostic due to Raftery and Lewis [61] and the convergence diagnostic test statistic (at a significance level of 5%) proposed by Geweke [36].

4.4.1 Quantile regression model with ALD errors

In the simulation study, we let the time related nonparametric part of the model in (4.1) be $g(t_{ij}) = \sin(2\pi t_{ij})$, such that $y_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + g(t_{ij}) + v_i + k_1 e_{ij} + r_{ij}\sqrt{k_2\theta e_{ij}},\ i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m$ with $m = 4$. The observations t_{ij}'s are generated from a uniform $[0, 1]$ distribution, $\mathbf{x}_{ij}$ is a 3×1 vector whose elements are independently sampled from a standard normal distribution $\mathrm{N}(0, 1)$ and $\boldsymbol{\beta} = (1, -0.8, 1)^T$. For the random effect v_i, we consider a log-linear structure of the variance model, such that

$\log(\sigma_i^2) = \mathbf{z}_i^T\boldsymbol{\gamma}$ with $\boldsymbol{\gamma} = (1, -0.5)^T$ and $\mathbf{z}_i = (z_{i1}, z_{i2})^T$, where z_{i1} and z_{i2} are independently sampled from $\mathrm{N}(0, 1)$. Then v_i is generated from the normal distribution $\mathrm{N}(0, \sigma_i^2)$. Since r_{ij} follows the normal distribution, $r_{ij}\sqrt{k_2\theta e_{ij}} \sim \mathrm{N}(0, k_2\theta e_{ij})$ with e_{ij} generated from the standard exponential distribution with $\theta = 1$, we generate y_{ij} from the normal distribution $\mathrm{N}(\mu_{ij}, k_2\theta e_{ij})$ with $\mu_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + g(t_{ij}) + v_i + k_1 e_{ij}$.

In practical application, one may argue that the specifications of the hyperparameters in the prior distributions could have a large impact on the posterior distributions of the parameters of interests. We here investigate the sensitivity of the proposed Bayesian procedure with three different types of the hyperparameter values for $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ as follows:

Type I: Accurate prior information with $\boldsymbol{\beta}_0 = (1, -0.8, 1)^T$ and $\boldsymbol{\gamma}_0 = (1, -0.5)^T$.

Type II: Inaccurate prior information with $\boldsymbol{\beta}_0 = 1.5 \times (1, -0.8, 1)^T$ and $\boldsymbol{\gamma}_0 = 1.5 \times (1, -0.5)^T$.

Type III: None prior information with $\boldsymbol{\beta}_0 = (0, 0, 0)^T$ and $\boldsymbol{\gamma}_0 = (0, 0)^T$.

Other hyperparameters are set as $\sigma_\gamma^2 = 4$, $a_\theta = b_\theta = 1$, $a_{\phi^2} = b_{\phi^2} = 1$, $\mathbf{B}_\beta = \mathbf{I}_3$, and $\mathbf{B}_\gamma = \mathbf{I}_2$. It deserves mentioning that the prior information can be easily included for the proposed procedure by specifying different values of the hyperparameters mentioned above. In addition, we also study the behavior of the proposed Bayesian approach with three different sample sizes $n = 30,\ 80,\ 160$ and quantile levels $\tau = 0.25,\ 0.5,\ 0.75$. We generate 100 replications from each combination of the above various settings.

To study the accuracy of estimating the nonparametric function $g(\cdot)$ based on the cubic B-spline approximation, we depict the true sine curve against its estimated one in Figure 4.1 and 4.2 under different scenarios. We observe from these figures that the B-spline method works very well for estimating nonparametric part of the model and that there is no observable effects of prior information among the above three different types of the hyperparameters, since all the estimated curves are close to the true sine curves. More simulation studies with respect to other values of the hyperparameters, n, and τ have also been conducted, and the conclusions are substantively similar and are thus not presented here for simplicity.

In Table 4.1, we present the simulation results for the unknown model parameters θ, β_1, β_2, β_3, γ_1, and γ_2 in terms of the estimated bias and the mean squared error (MSE) under different quantiles, sample size, and the three types of prior information. Some conclusions from this table can be summarized as follows:

1. As one expects, the bias and MSE of all the parameters decrease significantly under each quantile level as the sample size increases. For instance, there is no apparent difference between $n = 80$ and $n = 160$, which indicates that the

sample size $n = 80$ is large enough to obtain accurate estimates under the considered simulation settings.

2. The Bayesian estimates are quite robust for the specifications of the priors for the unknown parameters. The bias and MSE of the parameters do not have distinct difference across the three types of prior information. This indicates that the estimators converge to a certain level and the initial values of parameters in the proposed sampling algorithm do not affect the accuracy of the point estimations.

3. The parameter $\boldsymbol{\gamma}$ is related to the variance of the random effect and it has relatively large bias and MSE compared to other parameters. The hyperparameter θ always has a small bias regardless of different sample sizes and prior information.

4. At a specific combination of sample size and prior information, the biases under different quantiles are reasonably close to each other.

4.4.2 Quantile regression model with non ALD of errors

In this section, in order to check the performance of the proposed model under different data-generating error distributions, three different non-standard error distributions are used in the simulation. The semiparametric mixed-effects model is

$y_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + g(t_{ij}) + v_i + \varepsilon_{ij}$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$, where $m = 4$. Other simulation settings for generating observations and hyperparameters are the same as section 4.1. Sample size is set to be $n = 80$, three different quantile levels $\tau = 0.25,\ 0.5,\ 0.75$ are applied. For the prior information, we set $\boldsymbol{\beta}_0 = (0, 0, 0)^T$ and $\boldsymbol{\gamma}_0 = (0, 0)^T$, which are noninformative priors.

Data are generated under the following three distributions for the error term:

Type A : $\varepsilon_{ij} \sim \mathrm{N}(\mu, 4)$, with μ chosen such that the τth quantile is 0.

Type B : $\varepsilon_{ij} \sim \mathrm{Laplace}(\mu, 2)$, with μ chosen such that the τth quantile is 0.

Type C : $\varepsilon_{ij} \sim 0.3\mathrm{N}(\mu + 1, 1) + 0.7\mathrm{N}(\mu, 4)$, with μ chosen such that the τth quantile is 0.

By comparing Table 4.2 with the block of Type III in Table 4.1, the difference is negligible. Under the mixture normal distribution of error, the MSE's are smaller for all the parameters, particularly for 0.25 and 0.75 quantiles. One noteworthy feature is that the mixture normal distribution has a much better estimate than the ones under other distributions for the variance parameter $\boldsymbol{\gamma}$ when n is small (e.g., $n = 30$). Overall, the proposed model is relatively robust to the non-ALD of errors.

4.5 Real data application: the multi-center AIDS cohort study

In this section, we apply our Bayesian quantile mixed-effects model to the widely used MultiCenter AIDS Cohort Study (MACS) data. MACS is an ongoing study of HIV infection among homosexual men, initialed in 1984 at four institutions: UCLA, Northwestern university in Chicago, the university of Pittsburgh, and Johns Hopkins university in Baltimore. The latest dataset involves more than 7,000 gay men. In this paper, the data is collected from 1984 to 1991, containing 283 HIV positive gay men. This dataset has been widely used to study the mean CD4 percentage depletion over time and the effects of other physical status, including age of the patient at the start of the trial, smoking status and the post-infection CD4 percentage. Since the trend of CD4 depletion may be very different between high CD4 percentage patients and low CD4 percentage patients, this motivates us to apply a quantile regression model to investigate such groups of patients.

In the following Bayesian quantile regression model, y_{ij} is the observation of CD4 percentage at the current visit, x_{ij1} is the smoking status (0 for non-smoker and 1 for smoker), x_{ij2} is the age of the patient at the start of the trial and x_{ij3} is the post-infection CD4 percentage. To eliminate the intercept, $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})^T$ has been centered. $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ is the parameters for the linear part of the model. We assume the variance of the random effect v_i has a linear relationship with two explanatory variables $z_{i1} = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij1}$ and $z_{i2} = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij3}$. τ is the quantile such that $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $k_2 = \frac{2}{\tau(1-\tau)}$.

$$
\begin{cases}
y_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + g(t_{ij}) + v_i + k_1 e_{ij} + r_{ij}\sqrt{k_2\theta e_{ij}} \\
e_{ij} \sim \mathrm{Exp}(\theta^{-1}) \\
r_{ij} \sim \mathrm{N}(0,1) \\
v_i \sim \mathrm{N}(0, \sigma_i^2) \\
\sigma_i^2 = \gamma_1 z_{i1} + \gamma_2 z_{i2} \\
i = 1, 2, \cdots, 283, \quad j = 1, 2, \cdots, m_i
\end{cases}
$$

The main objective of this study is to figure out the relationship between mean CD4 percentage $g(t_{ij})$ and time t_{ij} at different quantiles. The Gibbs sampling algorithm is used with 55,000 iterations. All hyperparameters are set to be small and the initial value of all unknown parameters are noninformative of 0 value. In the Metropolis-Hastings procedure of sampling $\boldsymbol{\gamma}$, σ_γ^2 is set to 4, such that approximate acceptance rates for $\boldsymbol{\gamma}$ is 35%. 5 different quantiles are engaged in our analysis, which are 5%, lower quantile, median, upper quantile and 95%. The Bayesian estimates (BST) of θ, ϕ^2, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and their corresponding standard deviation estimates (SD) are summarized in the Table 4.3. All parameters have small standard deviations.

Figure 4.3 represents the estimated CD4 depletion trends under 5 different quantiles, the scattered data points are also plotted with the 5 curves. As predicted, the CD4 depletion trends are indeed distinct among diverse groups of patients. For the patients who have a high CD4 level around 50, their CD4 percentage almost get back to their original level after 6 years. The other patients' CD4 percentage decreases quickly after infection, however it increases a little bit at the end. This indicates drug usage can be based on the patients' starting CD4 percentage.

4.6 Discussion

In this study, we have developed the Bayesian quantile regression for semiparametric mixed effect model. The proposed model is first checked by the simulation model consisting of both linear and nonlinear part. The model has a good performance for estimating the linear parameters and fitting the sine curves. While for the parameters in the variance model, it is somewhat sensitive to the sample size, the estimated variance parameters have relative large bias and MSE when the sample size is small. For all other parameters, the model is quite robust to the prior information and sample size.

Eventually, the proposed model is applied to the famous MACS data, which has been widely analyzed by many authors. Most authors have adopted mean model for this dataset, such as Fan and Li [30] and Zhao and Xue [75]. The median model in this paper can be compared with their mean model. The middle curve in Figure 4.3 is indeed similar to Zhao and Xue [75]'s result.

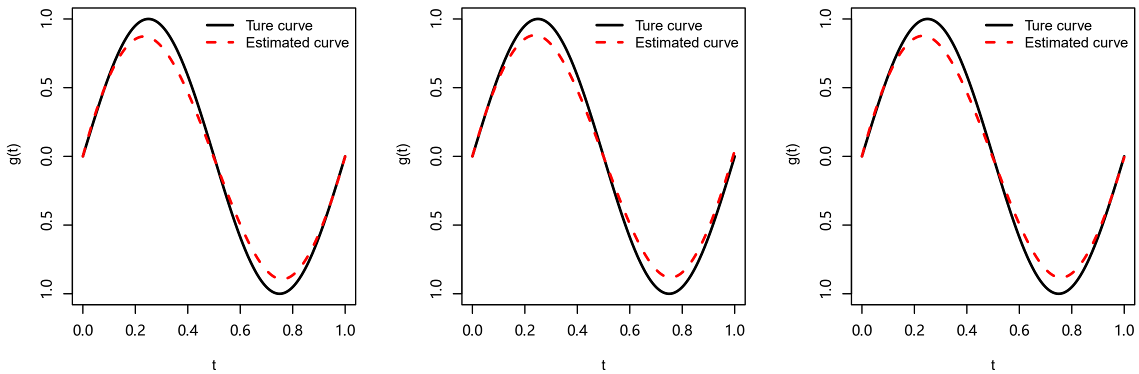


Figure 4.1: The true sine curve versus the estimated curve when $n = 80$ and quantile $\tau = 0.5$.
Type I (left panel), Type II (middle panel), Type III (right panel)

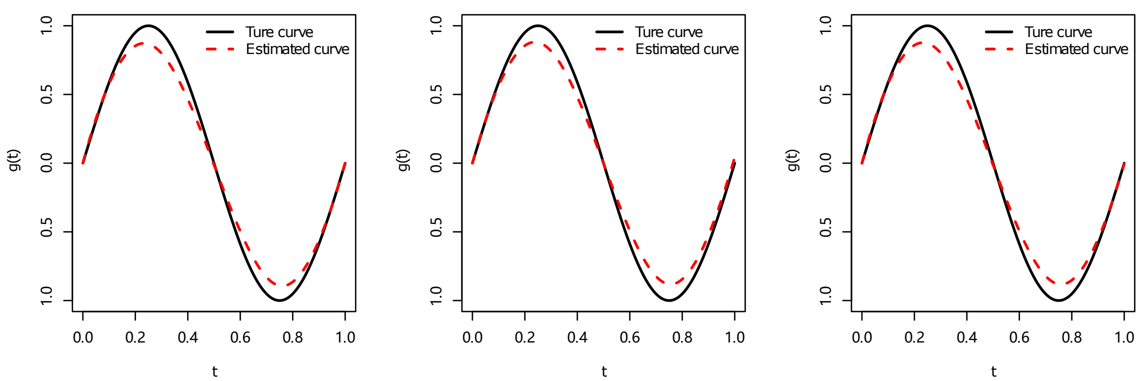


Figure 4.2: The true sine curve versus the estimated curve when $n = 160$ and quantile $\tau = 0.25$.
Type I (left panel), Type II (middle panel), Type III (right panel)

Table 4.1
BIAS and MSE (in parenthesis) of the Bayesian estimated parameters under AL error distributions.

		$n=30$			$n=80$			$n=160$		
Type	Par	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$	$\tau=0.25$	$\tau=0.5$	$\tau=0.75$
I	θ	−0.0120	−0.0052	0.0070	−0.0066	−0.0058	0.0047	−0.0027	0.0014	−0.0060
		(0.1009)	(0.0954)	(0.0997)	(0.0681)	(0.0554)	(0.0584)	(0.0406)	(0.0469)	(0.0388)
	β_1	−0.0292	0.0389	−0.0033	0.0186	0.0005	0.0055	−0.0110	−0.0097	−0.0110
		(0.0670)	(0.0509)	(0.0671)	(0.0671)	(0.0196)	(0.0267)	(0.0126)	(0.0081)	(0.0155)
	β_2	−0.0264	−0.0182	−0.0261	0.0157	−0.0177	0.0090	−0.0034	−0.0038	−0.0009
		(0.0926)	(0.0493)	(0.0850)	(0.0182)	(0.0182)	(0.0267)	(0.0122)	(0.0075)	(0.0146)
	β_3	0.0442	0.0383	−0.0158	0.0139	0.0201	−0.0154	0.0139	0.0152	0.0177
		(0.0631)	(0.0479)	(0.0610)	(0.0269)	(0.0244)	(0.0324)	(0.0119)	(0.0092)	(0.0122)
	γ_1	−0.0893	−0.0942	−0.0900	−0.0364	−0.0408	−0.1162	−0.0644	−0.0065	−0.0213
		(0.1575)	(0.1454)	(0.1915)	(0.1024)	(0.0749)	(0.1284)	(0.0559)	(0.0339)	(0.0499)
	γ_2	0.0015	0.0313	0.0981	0.0207	0.0226	0.0358	0.0593	0.0254	0.0139
		(0.1468)	(0.1433)	(0.1753)	(0.0839)	(0.0688)	(0.1168)	(0.0772)	(0.0520)	(0.0589)
II	θ	−0.0087	−0.0030	0.0059	−0.0038	−0.0026	−0.0024	0.0026	−0.0049	−0.0065
		(0.0840)	(0.1052)	(0.0897)	(0.0611)	(0.0629)	(0.0546)	(0.0400)	(0.0425)	(0.0433)
	β_1	0.0265	0.0484	0.0432	−0.0080	0.0228	−0.0007	−0.0113	0.0187	−0.0015
		(0.0635)	(0.0505)	(0.0570)	(0.0312)	(0.0201)	(0.0176)	(0.0129)	(0.0113)	(0.0123)
	β_2	−0.0411	0.0484	0.0040	−0.0012	−0.0085	−0.0123	0.0159	−0.0079	−0.0195
		(0.0693)	(0.0693)	(0.0655)	(0.0287)	(0.0199)	(0.0282)	(0.0159)	(0.0080)	(0.0148)
	β_3	0.0406	0.0220	0.0307	−0.0085	0.0036	0.0291	0.0025	0.0005	0.0188
		(0.0841)	(0.0507)	(0.0676)	(0.0206)	(0.0181)	(0.0293)	(0.0134)	(0.0089)	(0.0089)
	γ_1	0.0621	0.0383	0.0893	0.0396	−0.0371	−0.0476	−0.0756	0.0191	−0.0602
		(0.1373)	(0.1458)	(0.1301)	(0.1004)	(0.1096)	(0.0876)	(0.0733)	(0.0438)	(0.0565)
	γ_2	−0.0817	0.0101	−0.0500	0.0275	−0.0073	0.0253	0.0254	0.0124	0.0015
		(0.1657)	(0.1585)	(0.1790)	(0.1043)	(0.0946)	(0.1285)	(0.0546)	(0.0513)	(0.0646)
III	θ	0.0133	0.0054	0.0052	0.0066	−0.0017	0.0070	0.0082	−0.0029	0.0008
		(0.0902)	(0.0954)	(0.0875)	(0.0615)	(0.0555)	(0.0589)	(0.0385)	(0.0423)	(0.0438)
	β_1	−0.0534	−0.0554	−0.1049	−0.0060	−0.0137	−0.0254	−0.0077	−0.0179	−0.0129
		(0.0661)	(0.0460)	(0.0911)	(0.0209)	(0.0224)	(0.0256)	(0.0131)	(0.0087)	(0.0107)
	β_2	0.0289	0.0295	0.0555	0.0096	0.0385	0.0148	0.0118	0.0079	0.0185
		(0.0600)	(0.0457)	(0.0641)	(0.0295)	(0.0196)	(0.0175)	(0.0116)	(0.0103)	(0.0107)
	β_3	−0.0658	−0.0488	−0.1025	−0.0290	−0.0215	−0.0144	−0.0238	−0.0240	−0.0113
		(0.0620)	(0.0662)	(0.0936)	(0.0342)	(0.0233)	(0.0251)	(0.0156)	(0.0112)	(0.0107)
	γ_1	−0.3904	−0.3864	−0.4605	−0.2155	−0.1535	−0.1697	−0.0769	−0.0903	−0.1054
		(0.3888)	(0.3222)	(0.4177)	(0.1666)	(0.1105)	(0.1436)	(0.0640)	(0.0470)	(0.0619)
	γ_2	0.2749	0.2181	0.2131	0.0822	0.0510	0.0975	0.0309	0.0627	0.0771
		(0.0619)	(0.2544)	(0.2342)	(0.1289)	(0.0997)	(0.1173)	(0.0548)	(0.0548)	(0.0651)

Table 4.2

BIAS and MSE (in parenthesis) of the Bayesian estimated parameters under non-AL error distributions.

Type	Par	$n=30$ $\tau=0.25$	$n=30$ $\tau=0.5$	$n=30$ $\tau=0.75$	$n=80$ $\tau=0.25$	$n=80$ $\tau=0.5$	$n=80$ $\tau=0.75$	$n=160$ $\tau=0.25$	$n=160$ $\tau=0.5$	$n=160$ $\tau=0.75$
A	β_1	−0.0207	−0.0271	−0.0327	−0.0361	−0.0098	−0.0328	−0.0115	0.0019	−0.0274
		(0.0388)	(0.0464)	(0.0469)	(0.0190)	(0.0154)	(0.0181)	(0.0092)	(0.0071)	(0.0109)
	β_2	0.0314	0.0334	0.0319	0.0124	0.0136	0.0274	0.0012	0.0024	0.0105
		(0.0481)	(0.0333)	(0.0537)	(0.0139)	(0.0157)	(0.0176)	(0.0117)	(0.0083)	(0.0083)
	β_3	−0.0390	−0.1331	−0.0460	−0.0201	−0.0200	−0.0246	−0.0074	−0.0181	−0.0111
		(0.0462)	(0.0549)	(0.0619)	(0.0184)	(0.0159)	(0.0142)	(0.0096)	(0.0084)	(0.0115)
	γ_1	−0.3511	−0.3409	−0.3111	−0.1329	−0.1749	−0.1490	−0.1109	−0.0847	−0.0570
		(0.3303)	(0.3538)	(0.3077)	(0.1027)	(0.0985)	(0.0859)	(0.0496)	(0.0409)	(0.0528)
	γ_2	0.1616	0.1717	0.1973	0.1402	0.0446	0.0677	0.0557	0.0144	−0.0239
		(0.2122)	(0.1398)	(0.2131)	(0.0819)	(0.0736)	(0.0672)	(0.0587)	(0.0386)	(0.0573)
B	β_1	−0.0546	−0.0319	−0.0635	−0.0144	−0.0002	−0.0141	−0.0196	−0.0136	−0.0075
		(0.0792)	(0.0580)	(0.0706)	(0.0216)	(0.0208)	(0.0309)	(0.0153)	(0.0097)	(0.0125)
	β_2	0.0478	0.0264	0.0322	0.0373	0.0178	−0.0082	0.0071	0.0136	0.0273
		(0.0874)	(0.0447)	(0.0715)	(0.0231)	(0.0154)	(0.0281)	(0.0154)	(0.0090)	(0.0135)
	β_3	−0.0814	−0.0965	−0.1166	−0.0289	0.0005	−0.0160	−0.0097	−0.0058	−0.0091
		(0.0781)	(0.0876)	(0.1116)	(0.0206)	(0.0186)	(0.0259)	(0.0141)	(0.0125)	(0.0131)
	γ_1	−0.4020	−0.3837	−0.4462	−0.1167	−0.1514	−0.1454	−0.1393	−0.0552	−0.0896
		(0.4017)	(0.3499)	(0.3815)	(0.1201)	(0.1158)	(0.1099)	(0.1122)	(0.0441)	(0.0789)
	γ_2	0.2355	0.1080	0.1945	0.1040	0.1044	0.0913	0.0970	0.0391	0.0006
		(0.2878)	(0.1769)	(0.3284)	(0.1139)	(0.0915)	(0.1163)	(0.1020)	(0.0619)	(0.0716)
C	β_1	−0.0438	−0.0344	−0.0562	−0.0066	−0.0194	−0.0144	−0.0165	−0.0110	−0.0108
		(0.0308)	(0.0253)	(0.0270)	(0.0080)	(0.0066)	(0.0081)	(0.0053)	(0.0053)	(0.0038)
	β_2	0.0260	0.0415	0.0357	0.0226	0.0010	0.0100	0.0047	−0.0041	−0.0071
		(0.0223)	(0.0279)	(0.0312)	(0.0087)	(0.0083)	(0.0094)	(0.0055)	(0.0037)	(0.0047)
	β_3	−0.0767	−0.0432	−0.0615	−0.0045	0.0040	−0.0126	−0.0019	−0.0031	−0.0152
		(0.0309)	(0.0233)	(0.0245)	(0.0072)	(0.0064)	(0.0092)	(0.0066)	(0.0046)	(0.0054)
	γ_1	−0.1808	−0.2185	−0.1716	−0.0950	−0.0337	−0.1084	−0.0825	−0.0366	−0.0885
		(0.1779)	(0.1750)	(0.2245)	(0.0603)	(0.0497)	(0.0641)	(0.0344)	(0.0285)	(0.0383)
	γ_2	0.0914	0.0746	0.0587	0.0182	0.0400	0.0670	0.0260	0.0380	0.0475
		(0.1283)	(0.1305)	(0.1276)	(0.0420)	(0.0402)	(0.0534)	(0.0295)	(0.0326)	(0.0333)

Table 4.3

Bayesian estimates of parameters

Para -meter	$\tau=0.05$ EST	$\tau=0.05$ SD	$\tau=0.25$ EST	$\tau=0.25$ SD	$\tau=0.50$ EST	$\tau=0.50$ SD	$\tau=0.75$ EST	$\tau=0.75$ SD	$\tau=0.95$ EST	$\tau=0.95$ SD
θ	13.2085	0.4839	4.4710	0.1204	4.1962	0.0993	4.5024	0.1200	13.2236	0.4842
β_1	2.0488	1.6893	−0.7739	0.7027	0.2189	0.6922	1.5626	0.6835	−0.7623	1.7094
β_2	0.0355	0.1134	−0.1403	0.0463	−0.1980	0.0506	−0.1137	0.0436	0.0135	0.1074
β_3	0.1562	0.1046	0.4223	0.0454	0.4630	0.0464	0.4889	0.0384	0.4189	0.0988
γ_1	−0.0675	0.9112	−11.8339	0.3957	−12.1509	0.3400	−11.3330	0.3936	−0.1292	0.9174
γ_2	0.0351	0.1116	0.0265	0.0167	0.0161	0.0142	0.0085	0.0164	−0.0505	0.1049

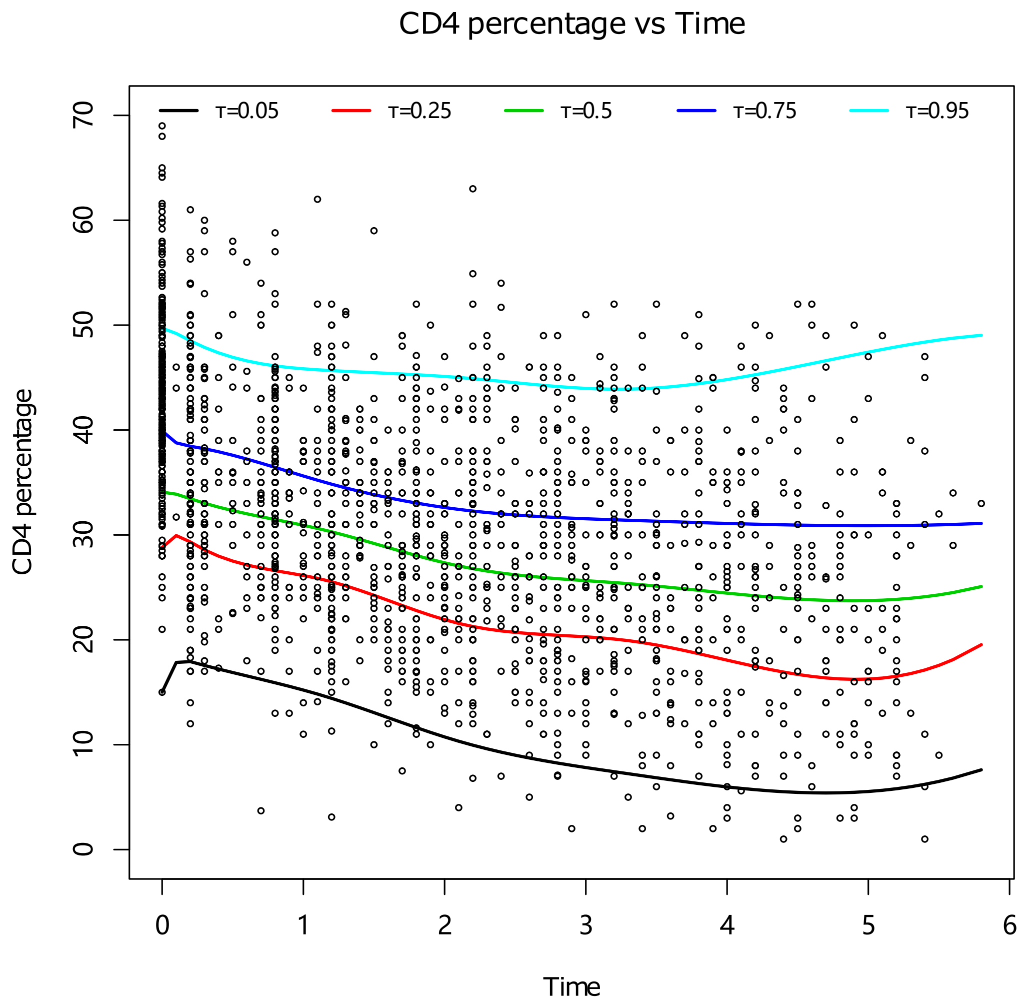


Figure 4.3: The mean CD4 percentage $g(t_{ij})$ vs time t_{ij} at different quantiles.

Chapter 5

Future Work

While the Bayesian hypothesis testing and point estimation problems of ICC have been widely studied, there are still loads of future studies that need to be conducted. In Chapter 2, we derive the Bayes factor based on the divergence-based priors in linear models. The proposed method could be extended to other statistical models, such as the network autocorrelation model discussed in the following section.

For the semiparametric models in Chapter 4, the parameters are estimated from their posterior distributions through Markov Chain Monte Carlo algorithm. It is of particular interest to conduct variable selection to identify some important predictors that are contributive to the response variable. Variable selection will not only significantly shorten training times, but also boost computational efficiency. This observation motivates us to investigate problem of variable section in Bayesian quantile regression for semiparametric mixed-effect models.

5.1 Bayes factor based on the divergence-based priors for the network autocorrelation model

The network autocorrelation model is expressed as

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.1}$$

where $\mathbf{y}$ is an $n \times 1$ $(k \geq 2)$ vector for values of a dependent variable, $\mathbf{X}$ is an $n \times k$ matrix of values for the n actors on k independent variables, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown regression coefficients. We assume that the random error $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$, where $\overset{iid}{\sim}$ stands for "independent and identically distributed", $\mathbf{0}_n$ is an $n \times 1$ vector of zeros and $\mathbf{I}_n$ being an $n \times n$ identity matrix. Furthermore, $\mathbf{W}$ is an $n \times n$ weight matrix representing social ties in a network, with w_{ij} denoting the degree to which y_i depends on y_j $(i, j = 1, 2, \cdots, n)$. The key parameter ρ is referred as the network autocorrelation, which quantifies the social influence for given $\mathbf{y}$, $\mathbf{W}$ and $\mathbf{X}$.

The frequentist approach for testing $H_1 : \rho = 0$ versus $H_2 : \rho \neq 0$ is the significance tests, such as the likelihood ratio test. The decision rule is to compare the p-value with a prespecified significance level α, whereas it may not be able to evaluate evidence in favor of the null hypothesis as the Bayesian approach. This observation motivates us to consider the hypothesis testing problem from a Bayesian perspective. Dittrich et al. [27] proposed the Bayes factors based on empirical informative prior and uniform prior and obtained some interesting results through simulation studies and real-data applications. We consider the Bayes factors based on the divergence-based (DB) priors for the network autocorrelation coefficient ρ.

We observe from Doreian [29] that the probability density function (pdf) of $\mathbf{y}$ is given by

$$f(\mathbf{y} \mid \rho, \sigma^2, \boldsymbol{\beta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \mid \det(\mathbf{A}_\rho) \mid \exp\left\{ -\frac{1}{2\sigma^2}(\mathbf{A}_\rho\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{A}_\rho\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\},$$

where $A_\rho = \mathbf{I}_n - \rho\mathbf{W}$. The condition of a nonsingular $\mathbf{A}_\rho$ is $\rho \in \left(\lambda_{(n)}^{-1}, \lambda_{(1)}^{-1}\right)$, where $\lambda_{(1)} \leq \lambda_{(2)} \leq \cdots \leq \lambda_{(n)}$ are the ordered eigenvalues of $\mathbf{W}$; see Hepple [42].

As illustrated in Section 2.2.1, derivation of the DB priors requires the parameters to be orthogonal. First we need to translate $(\rho, \boldsymbol{\beta}, \sigma^2)$ to be the orthogonal parameters $(\theta_1 = \rho, \theta_2, \theta_3)$ by letting the off-diagonal elements of the Fisher information matrix equaling 0.

The hypothesis testing can be equivalently expressed as the model selection problem of the two competing models given by

$$M_1 : f_1(\mathbf{y} \mid \boldsymbol{\nu}) = f(\mathbf{y} \mid 0, \boldsymbol{\nu}) \quad \text{versus} \quad M_2 : f_2(\mathbf{y} \mid \theta_1, \boldsymbol{\nu}) = f(\mathbf{y} \mid \theta_1, \boldsymbol{\nu}). \tag{5.2}$$

where $\boldsymbol{\nu} = (\theta_2, \theta_3)$. The Bayes factor (BF) in favor of M_2 and against M_1 is defined as

$$\mathrm{BF}_{21} = \frac{p(\mathbf{y} \mid M_2)}{p(\mathbf{y} \mid M_1)} = \frac{\int f_2(\mathbf{y} \mid \theta_1, \boldsymbol{\nu}) \pi^D(\theta_1, \boldsymbol{\nu})\, d\theta_1\, d\boldsymbol{\nu}}{\int f_1(\mathbf{y} \mid \boldsymbol{\nu}) \pi^N(\boldsymbol{\nu})\, d\boldsymbol{\nu}}, \tag{5.3}$$

where $\pi^N(\boldsymbol{\nu})$ is a noninformative prior and $\pi^D(\theta_1, \boldsymbol{\nu}) \propto \pi^D(\theta_1 \mid \boldsymbol{\nu}) \pi^N(\boldsymbol{\nu})$ with $\pi^D(\theta_1 \mid \boldsymbol{\nu})$ being the conditional DB prior.

In order to derive the DB priors, we first find out the KL divergence between the two models 5.2 and acquire the sum-DB measure and the min-DB measure, respectively. By following Definition 2.2.1, it is not difficult to derive the conditional DB prior $\pi^D(\theta_1 \mid \boldsymbol{\nu})$. Finally, by integrating $\boldsymbol{\nu}$ out in 5.3, we obtain the Bayes factors based on the sum and min-DB priors. In ongoing work, we compare the performance of the proposed Bayes factors under the DB priors and the ones based on empirical informative prior and uniform prior due to Doreian [29]. After extensive simulations studies, the results will be reported elsewhere.

5.2 Bayesian variable selection for the semiparametric mixed-effect models

Since Yu and Moyeed [70] proposed Bayesian quantile regression by employing the asymmetric Laplace distribution (ALD) for the error term, many researchers studied variable selection for Bayesian quantile regression, such as Koenker and Machado [49], Yu et al. [69], Yuan and Lin [71], and Alhamzawi and Yu [3]. More recently, Zhang et al. [74] developed Bayesian variable selection methods in semi-parametric models in the framework of partially linear Gaussian and problit regressions. We observe that most of Bayesian procedures for variable selection in quantile regression models consider the specification of priors independent of quantiles, even though the parameter values could vary with quantiles under consideration. This observation motivates us to develop a quantile dependent prior for regression coefficients that is as informative as possible. In ongoing work, we plan to develop a quantile dependent prior for the regression coefficients and conduct the problem of Bayesian variable selection in semiparametric mixed-effects double regression models.

For the variable selection problem, we could utilize indicator variables for variable inclusion and elimination, see Smith and Kohn [63], Kuo and Mallick [54] and Liang et al. [56] for inference. Other Bayesian variable selection methods including stochastic search variable selection (SSVS) (Yi et al. [68] and Brown et al. [18]), reversible jump MCMC (Green [39]) and composite model space (Godsill [38] and Fang et al. [31]) could be considered as well. These will be investigated in the future.

References

[1] Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 36(3), 332–339.

[2] Alhamzawi, R. and Ali, H. T. (2017). Bayesian quantile regression for ordinal longitudinal data. *Journal of Applied Statistics*, 45(5), 815–828.

[3] Alhamzawi, R. and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, 39(4), 799–813.

[4] Barkto, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3–11.

[5] Bayarri, M. J. and García-Donato, G. (2008). Generalization of Jeffreys divergence-based priors for Bayesian hypothesis testing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70(5), 981–1003.

[6] Berger, J. O. and Bernardo, J. M. (1991). On the development of reference priors. In *Bayesian Statistics, 4 (pp.35–60)*. New York: Oxford University Press.

[7] Berger, J. O. and Bernardo, J. M. (1992). Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, volume 75 of *Lecture Notes in Statistics (pp.177–194)*. New York: Springer.

[8] Berger, J. O., and Pericchi, L.R. (2001). Objective Bayesian methods for model selection: introduction and comparison. In *Model selection*, volume 38 of *IMS Lecture Notes Monograph Series (pp.135–207)*. Institute of Mathematical Statistics, Beachwood, OH. With discussion by Ghosh, J., Samanta, T., De Santis, F. and a rejoinder by the authors.

[9] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Reference posterior distributions for Bayesian inference.*, 41(2), 113–147.

[10] Bernardo, J. M. (1999). Nested hypothesis testing: the Bayesian reference criterion. In *Bayesian Statistics, 6 (pp.101–130)*. New York: Oxford University Press.

[11] Bernardo, J. M. (2005). Reference analysis. In *Bayesian Thinking: Modeling and Computation (Vol. 25, pp.17–90)*. Amsterdam, North-Holland: Elsevier.

[12] Bernardo, J. M. (2010). Integrated objective Bayesian estimation and hypothesis testing. In *Bayesian Statistics, 9. Proceedings of the Ninth Valencia International Meeting (pp. 1–68)*. New York: Oxford University Press.

[13] Bernardo, J. M. and Juárez, M. A. (2003). Intrinsic estimation. In *Bayesian Statistics, 7 (pp.465–476)*. New York: Oxford University Press.

[14] Bernardo, J. M. and Pérez, S. (2007). Comparing normal means: new methods for an old problem. In *Bayesian Analysis, 8, (pp.571-576.)*. New York: Oxford University Press.

[15] Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: a reference approach. *International Statistical Review*, 70(3), 351–372.

[16] Bernardo, J. M. and Smith, A. F. (1994). *Bayesian Theory.* Chichester: John Wiley.

[17] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

[18] Brown, P. J., Vannucci, M., and Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 60(3), 627-641.

[19] Cepeda, E. and Gamerman, D. (2000). Bayesian modeling of variance heterogeneity in normal regression models. *Brazilian Journal of Probability and Statistics*, 14(2), 207–221.

[20] Chen, C. and Yu, K. (2009). Automatic Bayesian quantile regression curve fitting. *Statistics and Computing*, 19(3), 271–281.

[21] Chib, S. and Greenberg, G. (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2), 183–206.

[22] Chung, Y. and Dey, D. K. (1998). Bayesian approach to estimation of intraclass correlation using reference prior. *Communications in Statistics - Theory and Methods*, 27(9), 2241–2255.

[23] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 49(1), 1–18.

[24] Datta, G. S. and Ghosh, J. K. (1995). Noninformative priors for maximal invariant parameter in group models. *TEST*, 4(1), 95–114.

[25] Datta, G. S. and Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, 82(1), 37–45.

[26] Datta, G. S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics Lecture Notes in Statistics*, volume 178 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

[27] Dittrich, D., Leenders, R. T. and Mulder, J. (2017). Network autocorrelation modeling: a Bayes factor approach for testing (multiple) precise and interval hypotheses. *Sociological Methods & Research*.

[28] Donner, A. (1979). The use of correlation and regression in the analysis of family resemblance. *American Journal of Epidemiology*, 110(3), 335–342.

[29] Doreian, P. (1980). Linear models with spatially distributed data: spatial disturbances or spatial effects? *Sociological Methods & Research*, 9(1), 29–60.

[30] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467), 710–723.

[31] Fang, M., Liu, S., and Jiang, D. (2009). Bayesian composite model space approach for mapping quantitative trait loci in variance component model. *Behavior Genetics*, 39(3), 337–46.

[32] Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiments.* New York: Wiley.

[33] Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences.* Cambridge, UK: Cambridge University Press.

[34] García-Donato, G. and Sun, D. (2007). Objective priors for hypothesis testing in one-way random effects models. *Canadian Journal of Statistics*, 35(2), 303–320.

[35] Gelman, A., Roberts, G. O. and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics, 5 (pp.599–607).* New York: Oxford University Press.

[36] Geweke, J (2011). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics, 4 (pp.169-193).* New York: Oxford University Press.

[37] Ghosh, M. and Heo, J. (2003). Noninformative priors, credible sets and Bayesian hypothesis testing for the intraclass model. *Journal of Statistical Planning and Inference*, 112(1-2), 133–146.

[38] Godsill, S. J. (2001). On the relationship between Markov chain Monte Carlo methods for model uncertainty. *Journal of Computational and Graphical Statistics*, 10(2), 230–248.

[39] Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732.

[40] Haario, H., Saksman, E. and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14(3), 375.

[41] He, X., Fung, W. K. and Zhu, Z. Y. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 100(472), 1176-1184.

[42] Hepple, L. W. (1995). Bayesian techniques in spatial and network econometrics: 1. model comparison and posterior odds. *Environment and Planning A: Economy and Space*, 27(3), 447-469.

[43] Jeffreys, H. (1961). *Theory of Probability.* 3rd edition, Oxford: Clarendon Press.

[44] Jelenkowska, T. H. (1998). Bayesian estimation of the intraclass correlation coefficients in the mixed linear model. *Applied Mathematics*, 43(2), 103–110.

[45] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

[46] Kass, R. E. and Vaidyanathan, S. K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 54(1), 129–144.

[47] Kim, Y., Lee, W. D., Kang, S. G. and Kim, D. H. (2017). Objective Bayesian hypothesis testing in regression models with first-order autoregressive residuals. *Communications in Statistics - Theory and Methods*, 46(12), 5872–5887.

[48] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.

[49] Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448), 1296–1310.

[50] Koenker, R. and Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71(1-2), 265–283.

[51] Kottas, A. and Krnjaji, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36(2), 297–319.

[52] Kotz, S., Kozubowski, T. J. and Podgorski, K. (2001). Classical symmetric Laplace distribution. In *The Laplace Distribution and Generalizations (pp.15–131)*. Boston, MA: Birkhauser.

[53] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11), 1565–1578.

[54] Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *Sankhya: The Indian Journal of Statistics: Series B*, 60(1), 65–81.

[55] Lee, S. A. and Kim, D. H. (2006). Bayesian hypothesis testing for intraclass correlation coefficient. *The Korean Communications in Statistics*, 13(3), 551–566.

[56] Liang, F., Paulo, R., Molina, G., Clyde, M. and Berger, J. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 410-423.

[57] Lin, L., Hedayat, A. S., Sinha, B. and Yang, M. (2002). Statistical methods in assessing agreement: models, issues, and tools. *Journal of the American Statistical Association*, 97(457), 257–270.

[58] Lombardia, M. J. and Sperlich, S. (2007). Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70(5), 913–930.

[59] Paul, S. R. (1990). Maximum likelihood estimation of intraclass correlation in the analysis of familial data: estimating equation approach. *Biometrika*, 77(3), 549–555.

[60] Paul, S. R. (1996). Score tests for interclass correlation in familial data. *Biometrics*, 52(3), 955–963.

[61] Raftery, A. E and Lewis, S. M. (2011). One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7(4), 493-497.

[62] Robert, C. P. (1996). Intrinsic losses. *Theory and Decision*, 40(2), 191–214.

[63] Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75(2), 317-343.

[64] Sun, D. and Ye, K. (1996). Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika*, 83(1), 55–65.

[65] Wang, J. (2012). Bayesian quantile regression for parametric nonlinear mixed effects models. *Statistical Methods and Applications*, 21(3), 279-295.

[66] Wang, M. and Sun, X. (2014). Bayes factor consistency for one-way random effects model. *Communications in Statistics - Theory and Methods*, 43(23), 5072–5090.

[67] Xu, D., Zhang, Z. and Wu, L. (2016). Bayesian analysis for semiparametric mixed-effects double regression models. *Hacettepe Journal of Mathematics and Statistics*, 45(1), 279–296.

[68] Yi, N., George, V., and Allison, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, 164(3), 1129-1138.

[69] Yu, K., Chen, C. W., Reed, C. and Dunson, D. B (2013). Bayesian variable selection in quantile regression. *Statistics and Its Interface*, 6(2), 261–274.

[70] Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54(4), 437–447.

[71] Yuan, M. and Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472), 1215–1225.

[72] Zhang, D., He, D., Sun, X., Lu, T. and Wang, M. (2018). Objective Bayesian hypothesis testing and estimation for the intraclass model. *Statistical Theory and Related Fields*, 2(1), 37–47. doi: 10.1080/24754269.2018.1481586

[73] Zhang, D. and Wang, M. (2018). Objective Bayesian inference for the intraclass

correlation coefficient in linear models. *Statistics & Probability Letters*, 137, 292–296. doi:10.1016/j.spl.2018.02.004

[74] Zhang, H., Maity, A., Arshad, H., Holloway, J. and Karmaus, W. (2016). Variable selection in semi-parametric models. *Statistical Methods in Medical Research*, 25(4), 1736–1752.

[75] Zhao, P. and Xue, L. (2009). Variable selection for semiparametric varying coefficient partially linear models. *Statistics & Probability Letters*, 79(20), 2148-2157.

Appendix A

Proofs of Theorems from Chapter 2

A.1 Proof of Proposition 1

For $\mathbf{\Sigma} = (1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k}\mathbf{W}$, it is easy to show that $|\mathbf{\Sigma}| = 1$,

$$
\begin{aligned}
\operatorname{tr}(\mathbf{\Sigma}) &= (1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k}\operatorname{tr}(\mathbf{W}) \\
&= nk(1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k}, \quad \text{and} \\
\operatorname{tr}(\mathbf{\Sigma}^{-1}) &= (1-\theta_1)^{(k-1)/k}\big(1+(k-1)\theta_1\big)^{1/k} n \operatorname{tr}(\mathbf{V}^{-1}) \\
&= nk(1-\theta_1)^{-1/k}\big(1+(k-1)\theta_1\big)^{-(k-1)/k}\big(1+(k-2)\theta_1\big).
\end{aligned}
$$

We observe from model (2.7) that the pdf of $\mathbf{y}$ under M_1 is given by

$$f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3) = (2\pi)^{-\frac{kn}{2}} \mid \theta_2^{-1}\mathbf{I}_{nk}|^{-\frac{1}{2}}\exp\left\{ -\frac{\theta_2}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)\right\},$$

and that the pdf of $\mathbf{y}$ under M_2 is given by

$$f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) = (2\pi)^{-\frac{kn}{2}} \mid \theta_2^{-1}\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left\{ -\frac{\theta_2}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}_3)\right\}.$$

The direct KL divergence between two models M_1 and M_2, denoted by $\mathrm{KL}[\theta_0 : \theta_1]$, is given by

$$\begin{aligned}\mathrm{KL}[\theta_0 : \theta_1] &= \int \log \frac{f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)}{f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)} f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)d\mathbf{y}\\ &= -\frac{1}{2}\log \mid \boldsymbol{\Sigma} \mid -\frac{\theta_2}{2}\,\mathrm{tr}\left[(\boldsymbol{\Sigma}^{-1} - \mathbf{I}_{nk})(\theta_2^{-1}\boldsymbol{\Sigma})\right]\\ &= \frac{nk}{2}\Big[(1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k} - 1\Big].\end{aligned}$$

Similarly, the KL divergence between two models M_2 and M_1, denoted by $\mathrm{KL}[\theta_1 : \theta_0]$, is given by

$$\begin{aligned}\mathrm{KL}[\theta_1 : \theta_0] &= \int \log \frac{f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)}{f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)} f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)d\mathbf{y}\\ &= \frac{1}{2}\log \mid \boldsymbol{\Sigma} \mid -\frac{\theta_2}{2}\,\mathrm{tr}\left[(\mathbf{I}_{nk} - \boldsymbol{\Sigma}^{-1})(\theta_2^{-1}\mathbf{I}_{nk})\right]\\ &= \frac{nk}{2}\Big[(1-\theta_1)^{-1/k}\big(1+(k-1)\theta_1\big)^{-(k-1)/k}\big(1+(k-2)\theta_1\big) - 1\Big].\end{aligned}$$

The sum-DB measure in (2.11) is

$$
\begin{aligned}
D^S[\theta_0, \theta_1] =& \mathrm{KL}[\theta_0 : \theta_1] + \mathrm{KL}[\theta_1 : \theta_0] \\
=& \frac{nk}{2}\left[\frac{(1-\theta_1)^{-(k-1)/k}}{(1+(k-1)\theta_1)^{1/k}} + \frac{(1-\theta_1)^{-1/k}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{(k-1)/k}} - 2\right],
\end{aligned}
$$

and the min-DB measure in (2.12) is

$$
\begin{aligned}
D^M[\theta_0, \theta_1] =& 2 \times \min\{\mathrm{KL}[\theta_0 : \theta_1],\ \mathrm{KL}[\theta_1 : \theta_0] \\
=& 2 \times \Big[\mathrm{KL}[\theta_1 : \theta_0] I(\theta_1 \geq 0) + \mathrm{KL}[\theta_0 : \theta_1] I(\theta_1 < 0)\Big] \\
=& nk\Big[(1-\theta_1)^{-1/k}\big(1+(k-1)\theta_1\big)^{-(k-1)/k}\big(1+(k-2)\theta_1\big) I(\theta_1 \geq 0) \\
& \quad + (1-\theta_1)^{-(k-1)/k}\big(1+(k-1)\theta_1\big)^{-1/k} I(\theta_1 < 0) - 1\Big].
\end{aligned}
$$

Then, we follow Definition 2.2.1 to derive the conditional DB priors as follows:

Given that the number of data points is nk ($\mathbf{y}$ is $nk \times 1$), we set the effective sample size $n^* = nk$. For the sum-DB prior, when $k \geq 2$ and $q \geq 0$, it follows

$$
\begin{aligned}
c_S(q) =& \int_{\theta_1} (1 + \overline{D}^S[\theta_0, \theta_1])^{-q} \ \pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) d\theta_1 \\
=& \int_{\theta_1} (1 + D^S[\theta_0, \theta_1]/n^*)^{-q} \ (1 - \theta_1)^{-1} \big(1 + (k-1)\theta_1\big)^{-1} d\theta_1 \\
=& 2^q \int_{-\frac{1}{k-1}}^{1} \frac{(1-\theta_1)^{q(k-1)/k}}{(1+(k-1)\theta_1)^{-q/k}} \left[1 + \frac{(1-\theta_1)^{(k-2)/k}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{(k-2)/k}}\right]^{-q} d\theta_1 \\
\leq& \int_{-\frac{1}{k-1}}^{1} \frac{(1-\theta_1)^{q-q/k}}{(1+(k-1)\theta_1)^{-q/k}} d\theta_1 \\
\leq& \int_{-\frac{1}{k-1}}^{1} \big(1 + (k-1)\theta_1\big)^{q} d\theta_1 < \infty,
\end{aligned}
$$

thus, $\underline{q}^S = \inf\{q \geq 0 : c_S(q) < \infty\} = 0$, $q_*^S = \underline{q}^S + 2^{-1} = 2^{-1}$, which provides

$$
\begin{aligned}
\pi^S(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) =& c_S^{-1}(q_*^S)(1 + \overline{D}^S[\theta_0, \theta_1])^{-q_*^S} \pi^N(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) \\
=& \frac{\sqrt{2}}{c_S} \left[\frac{(1-\theta_1)^{1/k+1}}{(1+(k-1)\theta_1)^{1/k-2}} + \frac{(1-\theta_1)^{-1/k+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-1}}\right]^{-\frac{1}{2}},
\end{aligned}
$$

where

$$
\begin{aligned}
c_S =& c_S(q_*^S) \\
=& \sqrt{2} \int_{-\frac{1}{k-1}}^{1} \left[\frac{(1-\theta_1)^{1/k+1}}{(1+(k-1)\theta_1)^{1/k-2}} + \frac{(1-\theta_1)^{-1/k+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-1}}\right]^{-\frac{1}{2}} d\theta_1.
\end{aligned}
$$

For the conditional min-DB prior, when $k \geq 2$ and $q \geq 0$, it follows

$$\begin{aligned}
c_M(q) &= \int_{\theta_1} (1 + \overline{D}^M[\theta_0, \theta_1])^{-q} \ \pi^N(\theta_1|\theta_2, \boldsymbol{\theta}_3) d\theta_1 \\
&= \int_{\theta_1} (1 + D^M[\theta_0, \theta_1]/n^*)^{-q} \ (1 - \theta_1)^{-1} \big(1 + (k-1)\theta_1\big)^{-1} d\theta_1 \\
&= \int_{-\frac{1}{k-1}}^{1} \Big[(1 - \theta_1)^{q/k-1} \big(1 + (k-1)\theta_1\big)^{-q/k+q-1} \big(1 + (k-2)\theta_1\big)^{-q} I(\theta_1 \geq 0) \\
&\qquad\qquad + (1 - \theta_1)^{-q/k+q-1} \big(1 + (k-1)\theta_1\big)^{q/k-1} I(\theta_1 < 0) \Big] d\theta_1.
\end{aligned}$$

Now we prove $c_M(q) < \infty$ if $q > 0$. When $\theta_1 \in (-(k-1)^{-1}, 0)$, we have

$$\begin{aligned}
c_M(q) &= \int_{-\frac{1}{k-1}}^{0} (1 - \theta_1)^{-q/k+q-1} \big(1 + (k-1)\theta_1\big)^{q/k-1} d\theta_1 \\
&\leq \int_{-\frac{1}{k-1}}^{0} C_1 \big(1 + (k-1)\theta_1\big)^{q/k-1} d\theta_1 < \infty \quad (C_1 \text{ is a constant}),
\end{aligned}$$

since the function $f(\theta_1) = (1 - \theta_1)^{-q/k+q-1}$ is continuous on $[-(k-1)^{-1}, 0]$, it has an upper bound (C_1). Thus, the integral is finite if $q/k - 1 > -1$.

When $\theta_1 \in [0, 1)$, we have

$$\begin{aligned}
c_M(q) &= \int_{0}^{1} (1 - \theta_1)^{q/k-1} \big(1 + (k-1)\theta_1\big)^{-q/k+q-1} \big(1 + (k-2)\theta_1\big)^{-q} d\theta_1 \\
&\leq \int_{0}^{1} C_2 (1 - \theta_1)^{q/k-1} d\theta_1 < \infty \quad (C_2 \text{ is a constant}),
\end{aligned}$$

since the function $f(\theta_1) = \big(1 + (k-1)\theta_1\big)^{-q/k+q-1} \big(1 + (k-2)\theta_1\big)^{-q}$ is continuous on

$[0,1]$, it has an upper bound (C_2). Thus, the integral is finite if $q/k-1>-1$.

Consequently, $\underline{q}^M = \inf\{q>0 : c_M(q)<\infty\} = 0$, $q_*^M = \underline{q}^M + 2^{-1} = 2^{-1}$, and

$$\begin{aligned}
\pi^M(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) =& c_M^{-1}(q_*^M)(1+\overline{D}^M[\theta_0,\theta_1])^{-q_*^M}\pi^N(\theta_1|\theta_2,\boldsymbol{\theta}_3)\\
=& c_M^{-1}\Big[(1-\theta_1)^{\frac{1-2k}{2k}}\big(1+(k-1)\theta_1\big)^{-\frac{k+1}{2k}}\big(1+(k-2)\theta_1\big)^{-\frac{1}{2}}I(\theta_1\geq 0)\\
&+(1-\theta_1)^{-\frac{k+1}{2k}}\big(1+(k-1)\theta_1\big)^{\frac{1-2k}{2k}}I(\theta_1<0)\Big],
\end{aligned}$$

where

$$\begin{aligned}
c_M = c_M(q_*^M) = \int_{-\frac{1}{k-1}}^{1}&\Big[(1-\theta_1)^{\frac{1-2k}{2k}}\big(1+(k-1)\theta_1\big)^{-\frac{k+1}{2k}}\big(1+(k-2)\theta_1\big)^{-\frac{1}{2}}I(\theta_1\geq 0)\\
&+(1-\theta_1)^{-\frac{k+1}{2k}}\big(1+(k-1)\theta_1\big)^{\frac{1-2k}{2k}}I(\theta_1<0)\Big]d\theta_1,
\end{aligned}$$

leading to the proof of Proposition 1.

A.2 Proof of Theorem 1

We observe from equation in (2.15) that the BF associated with the sum-DB prior can be written as

$$\mathrm{BF}_{21}^S = \frac{\iiint f_2(\mathbf{y}\mid\theta_1,\theta_2,\boldsymbol{\theta}_3)\pi^S(\theta_1\mid\theta_2,\boldsymbol{\theta}_3)\pi^S(\theta_2,\boldsymbol{\theta}_3)\,d\theta_1\,d\theta_2\,d\boldsymbol{\theta}_3}{\iint f_1(\mathbf{y}\mid\theta_0,\theta_2,\boldsymbol{\theta}_3)\pi^S(\theta_2,\boldsymbol{\theta}_3)\,d\theta_2\,d\boldsymbol{\theta}_3} = \frac{m_2^S(\mathbf{y})}{m_1^S(\mathbf{y})},$$

where $f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)$ and $f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)$ are the pdf's of two competing models in (2.7), $\pi^S(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1}$ and $\pi^S(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)$ is the conditional sum-DB prior in Proposition 1. Thus, the denominator part is given by

$$
\begin{aligned}
m_1^S(\mathbf{y}) =& \iint f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3) \pi^S(\theta_2, \boldsymbol{\theta}_3)\, d\theta_2\, d\boldsymbol{\theta}_3 \\
=& \pi^{\frac{p-nk}{2}} \Gamma\Big(\frac{nk-p}{2}\Big) \mid \mathbf{X}^T\mathbf{X} \mid^{-\frac{1}{2}} \big[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\big]^{-\frac{nk-p}{2}},
\end{aligned}
$$

where $\mathbf{H}_1 = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. The numerator part is given by

$$
\begin{aligned}
m_2^S(\mathbf{y}) =& \iiint f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3) \pi^S(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3) \pi^S(\theta_2, \boldsymbol{\theta}_3)\, d\theta_1\, d\theta_2\, d\boldsymbol{\theta}_3 \\
=& \pi^{\frac{p-nk}{2}} \Gamma\Big(\frac{nk-p}{2}\Big) c_S^{-1} \int_{-\frac{1}{k-1}}^{1} h_S(\theta_1) d\theta_1,
\end{aligned}
$$

where c_S is defined in Proposition 1 and

$$
\begin{aligned}
h_S(\theta_1) =& \sqrt{2}\left[\frac{(1-\theta_1)^{1/k+n(k-1)+1}}{(1+(k-1)\theta_1)^{1/k-n-2}} + \frac{(1-\theta_1)^{-1/k+n(k-1)+2}(1+(k-2)\theta_1)}{(1+(k-1)\theta_1)^{-1/k-n-1}}\right]^{-\frac{1}{2}} \\
& \times |\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}|^{-\frac{1}{2}} \big[\mathbf{y}^T(\mathbf{W}^{-1} - \mathbf{H}_2)\mathbf{y}\big]^{-\frac{nk-p}{2}},
\end{aligned}
$$

with $\mathbf{H}_2 = \mathbf{W}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}$, $\mathbf{W} = \mathbf{I}_n \otimes \mathbf{V}$ and $\mathbf{V} = (1-\theta_1)\mathbf{I}_k + \theta_1 \mathbf{J}_k$. It can be easily shown that

$$
\mathrm{BF}_{21}^S = \frac{m_2^S(\mathbf{y})}{m_1^S(\mathbf{y})} = c_S^{-1} |\mathbf{X}^T\mathbf{X}|^{\frac{1}{2}} \big[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\big]^{\frac{nk-p}{2}} \int_{-\frac{1}{k-1}}^{1} h_S(\theta_1) d\theta_1.
$$

Similarly, the BF based on the min-DB prior is given by

$$\mathrm{BF}_{21}^{M} = \frac{\iiint f_2(\mathbf{y} \mid \theta_1, \theta_2, \boldsymbol{\theta}_3)\pi^M(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)\pi^M(\theta_2, \boldsymbol{\theta}_3)\, d\theta_1\, d\theta_2\, d\boldsymbol{\theta}_3}{\iint f_1(\mathbf{y} \mid \theta_0, \theta_2, \boldsymbol{\theta}_3)\pi^M(\theta_2, \boldsymbol{\theta}_3)\, d\theta_2\, d\boldsymbol{\theta}_3} = \frac{m_2^M(\mathbf{y})}{m_1^M(\mathbf{y})},$$

where $\pi^M(\theta_2, \boldsymbol{\theta}_3) \propto \theta_2^{-1}$ and $\pi^M(\theta_1 \mid \theta_2, \boldsymbol{\theta}_3)$ is the conditional min-DB prior in Proposition 1. Thus, the denominator part is given by

$$m_1^M(\mathbf{y}) = m_1^S(\mathbf{y}) = \pi^{\frac{p-nk}{2}}\Gamma\Big(\frac{nk-p}{2}\Big) \mid \mathbf{X}^T\mathbf{X} \mid^{-\frac{1}{2}} \big[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\big]^{-\frac{nk-p}{2}}.$$

Integrating with respect to $(\theta_1, \theta_2, \boldsymbol{\theta}_3)$, the numerator part is given by

$$m_2^M(\mathbf{y}) = \pi^{\frac{p-nk}{2}}\Gamma\Big(\frac{nk-p}{2}\Big)c_M^{-1}\int_{-\frac{1}{k-1}}^{1} h_M(\theta_1)d\theta_1,$$

where c_M is defined in Proposition 1 and

$$\begin{aligned} h_M(\theta_1) =&\Big[(1-\theta_1)^{\frac{1}{2k}+\frac{n-nk-2}{2}}\big(1+(k-1)\theta_1\big)^{-\frac{1}{2k}-\frac{n+1}{2}}\big(1+(k-2)\theta_1\big)^{-\frac{1}{2}}I(\theta_1 \geq 0) \\ &+ (1-\theta_1)^{-\frac{1}{2k}+\frac{n-nk-1}{2}}\big(1+(k-1)\theta_1\big)^{\frac{1}{2k}-\frac{n+2}{2}}I(\theta_1 < 0)\Big] \\ &\times |\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X}|^{-\frac{1}{2}}\big[\mathbf{y}^T(\mathbf{W}^{-1} - \mathbf{H}_2)\mathbf{y}\big]^{-\frac{nk-p}{2}}. \end{aligned}$$

Thus, simple algebra shows that the resulting BF is given by

$$\mathrm{BF}_{21}^{M} = \frac{m_2^M(\mathbf{y})}{m_1^M(\mathbf{y})} = c_M^{-1}|\mathbf{X}^T\mathbf{X}|^{\frac{1}{2}}\big[\mathbf{y}^T(\mathbf{I}_{nk} - \mathbf{H}_1)\mathbf{y}\big]^{\frac{nk-p}{2}}\int_{-\frac{1}{k-1}}^{1} h_M(\theta_1)d\theta_1.$$

This completed the proof of Theorem 1.

Appendix B

Proofs of Theorems from Chapter 3

In this appendix, we prove that the posterior distribution is proper under π_R in (3.5), since the case for π_J is exactly the same and thus omitted for simplicity. We first provide a very useful lemma, which plays an important role in determining the tail behavior of the key terms of the marginal posterior distribution $\pi_R(\rho \mid D)$.

Lemma 1 *The marginal posterior distribution $\pi_R(\rho \mid D)$ in (3.6) is a continuous function in $\big(-1/(k-1),\ 1\big)$ and their terms are such that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1-\rho)^{p/2})$ and $\mathbf{S}(\rho) = O((1-\rho)^{-1})$ as $\rho \to 1$, and such that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1+\rho(k-1))^{p/2})$ and $\mathbf{S}(\rho) = O((1+\rho(k-1))^{-1})$ as $\rho \to -1/(k-1)$.*

Proof. Direct inspection shows that $\pi_R(\rho \mid D)$ in (3.6) is a continuous function in $\big(-1/(k-1),\ 1\big)$. We consider the behavior of its two key terms as (i) $\rho \to 1$ and (ii) $\rho \to -1/(k-1)$.

(i) Let $\eta_1 = \rho/(1-\rho)$, which tends to infinity as $\rho \to 1$. Given that $\mathbf{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k = (1-\rho)\big[\mathbf{I}_k + \rho/(1-\rho)\mathbf{J}_k\big]$, we have

$$\mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\bigg(\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k\bigg).$$

Then it follows that

$$\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X} = \sum_{i=1}^{n}\mathbf{X}_i'\mathbf{\Sigma}^{-1}\mathbf{X}_i = (1-\rho)^{-1}\sum_{i=1}^{n}\bigg(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\bigg). \tag{B.1}$$

As $\eta_1 \to \infty$, we have

$$\bigg|\sum_{i=1}^{n}\bigg(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\bigg)\bigg| = O(1),$$

which show that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}| = O((1-\rho)^{-p})$, and thus

$$|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1-\rho)^{p/2}).$$

In addition, as $\eta_1 \to \infty$, we observe that each element of the inverse matrix in the right hand of Equation (B.1) becomes $O(1)$. With a little abuse of notation, as

$\eta_1 \to \infty$, we denote

$$\left[\sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1 \mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\right)\right]^{-1} = O(1),$$

which shows that $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O((1-\rho))$. Note also that $\mathbf{\Phi}^{-1} = \mathbf{I}_n \otimes \mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\mathbf{I}_n \otimes \big(\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k\big) = (1-\rho)^{-1}\mathbf{\Phi}_1^{-1}$, where

$$\mathbf{\Phi}_1^{-1} = \mathbf{I}_n \otimes \left(\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k\right) \to \mathbf{I}_n \otimes \left(\mathbf{I}_k - \frac{1}{k}\mathbf{J}_k\right),$$

as $\eta_1 \to \infty$. Also, $(\mathbf{X}'\mathbf{\Phi}_1^{-1}\mathbf{X})^{-1} = (1-\rho)^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O(1)$. Thus, as $\rho \to 1$, it follows

$$\begin{aligned}
\mathbf{S}(\rho) &= \mathbf{Y}'\big(\mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1}\big)\mathbf{Y} \\
&= \frac{1}{1-\rho}\mathbf{Y}'\big(\mathbf{\Phi}_1^{-1} - \mathbf{\Phi}_1^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}_1^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}_1^{-1}\big)\mathbf{Y} \\
&= O\big((1-\rho)^{-1}\big).
\end{aligned}$$

(ii) Let $\eta_2 = \rho/\big(1+\rho(k-1)\big)$, which tends to infinity as $\rho \to -1/(k-1)$. Given that

$$\mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\left(\mathbf{I}_k - \frac{\rho}{1+\rho(k-1)}\mathbf{J}_k\right) = (1-\rho)^{-1}\left(\mathbf{I}_k - \eta_2\mathbf{J}_k\right),$$

it follows that

$$\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X} = \sum_{i=1}^{n}\mathbf{X}_i'\mathbf{\Sigma}^{-1}\mathbf{X}_i = (1-\rho)^{-1}\sum_{i=1}^{n}\Big(\mathbf{X}_i'\mathbf{X}_i - \eta_2\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i\Big).$$

As $\eta_2 \to \infty$, we have $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}| = O(\eta_2^p)$, and thus

$$|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O(\eta_2^{-p/2}) = O\big((1+\rho(k-1))^{p/2}\big).$$

In addition, as $\eta_2 \to \infty$, we observe that $\big(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}\big)^{-1} = O(1)$ and that $\mathbf{\Phi}^{-1} = \mathbf{I}_n \otimes \mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\mathbf{I}_n \otimes \big(\mathbf{I}_k - \eta_2\mathbf{J}_k\big) = \eta_2\mathbf{\Phi}_2^{-1}$, where

$$\mathbf{\Phi}_2^{-1} = \frac{1}{1-\rho}\mathbf{I}_n \otimes \left(\mathbf{J}_k - \frac{1}{\eta_2}\mathbf{I}_k\right) \to \frac{k-1}{k}\mathbf{I}_n \otimes \mathbf{J}_k.$$

As $\eta_2 \to \infty$, we have $\big(\mathbf{X}'\mathbf{\Phi}_2^{-1}\mathbf{X}\big)^{-1} = \eta_2\big(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}\big)^{-1} = O(1)$, and thus

$$\begin{aligned}
\mathbf{S}(\rho) &= \mathbf{Y}'\big(\mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1}\big)\mathbf{Y} \\
&= \eta_2\mathbf{Y}'\big(\mathbf{\Phi}_2^{-1} - \mathbf{\Phi}_2^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}_2^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}_2^{-1}\big)\mathbf{Y} \\
&= O(\eta_2) = \big((1+\rho(k-1))^{-1}\big).
\end{aligned}$$

$\square$

B.1 Proof of Theorem 2

We now show that the posterior distribution under π_R is proper. Recall that the corresponding marginal posterior of ρ is given by

$$\pi_R(\rho \mid D) \propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1}|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2}. \quad \text{(B.2)}$$

Then the reference prior π_R leads to a proper posterior distribution if and only if

$$\int_{-1/(k-1)}^{1} \pi_R(\rho \mid D)\, d\rho < \infty.$$

By following Lemma 1, we observe that $\rho \to 1$, the tail behavior of $\pi_R(\rho \mid D)$ follows

$$\begin{aligned}\pi_R(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1}|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2} \\ &= O\big((1-\rho)^{n/2-1}\big),\end{aligned}$$

and that $\rho \to -1/(k-1)$, the tail behavior of $\pi_R(\rho \mid D)$ follows

$$\begin{aligned}\pi_R(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1}|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2} \\ &= O\big((1+\rho(k-1))^{n(k-1)/2-1}\big).\end{aligned}$$

Given that $\pi_R(\rho \mid D)$ is a continuous function in $\big(-1/(k-1),\ 1\big)$, the posterior distribution under π_R is proper, provided that $k \geq 2$. This completed the proof of Theorem 2.

B.2 Proof of Theorem 3

Define $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ and $\boldsymbol{\Sigma}_0 = (1-\rho_0)\mathbf{I}_k + \rho_0\mathbf{J}_k$. It can be easily verified that

$$
\begin{aligned}
\operatorname{tr}\big(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big) &= \frac{k\big(1+(k-2)\rho_0 - (k-1)\rho\rho_0\big)}{(1-\rho_0)(1+(k-1)\rho_0)} \\
\big|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big| &= \frac{\big(1+(k-1)\rho\big)(1-\rho)^{k-1}}{\big(1+(k-1)\rho_0\big)(1-\rho_0)^{k-1}},
\end{aligned}
$$

where $\operatorname{tr}(\mathbf{M})$ represents the trace of the matrix $\mathbf{M}$.

Consider that the KL divergence measure of a normal linear model $N_{kn}\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)\big)$ from another normal linear model $N_{kn}\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\big)$ is given by

$$
\begin{aligned}
&\int p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\big) \log \frac{p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\big)}{p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)\big)}\, d\mathbf{y} \\
&= \frac{1}{2}\left\{ \frac{R_0}{\sigma_0^2} + \mathrm{tr}\Big(\frac{\sigma^2}{\sigma_0^2}(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)^{-1}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\Big) - \log\Big|\frac{\sigma^2}{\sigma_0^2}(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)^{-1}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\Big| - kn \right\} \\
&= \frac{1}{2}\left\{ \frac{R_0}{\sigma_0^2} + \mathrm{tr}\Big(\frac{\sigma^2}{\sigma_0^2}\mathbf{I}_n \otimes (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\Big) - \log\Big|\frac{\sigma^2}{\sigma_0^2}\mathbf{I}_n \otimes (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\Big| - kn \right\} \\
&= \frac{1}{2}\left\{ \frac{R_0}{\sigma_0^2} + n\frac{\sigma^2}{\sigma_0^2}\mathrm{tr}\big(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big) - nk\log\Big(\frac{\sigma^2}{\sigma_0^2}\Big) - n\log\big|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big| - kn \right\},
\end{aligned}
$$

where $R_0 = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)^{-1}\mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$. The minimum of the logarithmic divergence above for $\boldsymbol{\beta}_0 \in R^p$ and $\sigma_0 > 0$ is achieved when

$$
\boldsymbol{\beta}_0 = \boldsymbol{\beta} \quad \text{and} \quad \sigma_0 = \sigma\sqrt{\frac{\mathrm{tr}\big(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big)}{k}},
$$

and substitution yields

$$
\begin{aligned}
&\kappa(\rho_0 \mid \sigma^2, \boldsymbol{\beta}, \rho) \\
&= \inf_{\boldsymbol{\beta}_0 \in R^p, \sigma_0 > 0} \frac{1}{2}\left\{ \frac{R_0}{\sigma_0^2} + n\frac{\sigma^2}{\sigma_0^2}\mathrm{tr}\big(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\big) - nk\log\Big(\frac{\sigma^2}{\sigma_0^2}\Big) - n\log\Big|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\Big| - kn \right\} \\
&= \frac{n}{2}\left\{ k\log\big(\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\big) - \log\big(|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}|\big) - k\log(k) \right\} \\
&= \frac{nk}{2}\log\left\{ \frac{1 + (k-2)\rho_0 - (k-1)\rho\rho_0}{(1 + (k-1)\rho_0)(1-\rho_0)} \right\} - \frac{n}{2}\log\left\{ \frac{(1 + (k-1)\rho)(1-\rho)^{k-1}}{(1 + (k-1)\rho_0)(1-\rho_0)^{k-1}} \right\},
\end{aligned}
$$

which is the same as $\kappa(\rho_0 \mid \rho)$ in (3.9).

Similarly, the minimum of the logarithmic divergence measure of $N_{kn}\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\big)$ from $N_{kn}\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)\big)$ is given by

$$\begin{aligned}
&\int p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)\big) \log \frac{p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)\big)}{p\big(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\big)}\, d\mathbf{y} \\
&= \frac{1}{2}\left\{ \frac{R}{\sigma^2} + n\frac{\sigma_0^2}{\sigma^2}\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\big) - nk\log\Big(\frac{\sigma_0^2}{\sigma^2}\Big) - n\log\big|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\big| - kn \right\},
\end{aligned}$$

where $R = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{I}_n \otimes \boldsymbol{\Sigma})^{-1}\mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$. The minimum of the divergence measure above for $\boldsymbol{\beta}_0 \in R^p$ and $\sigma_0 > 0$ is achieved when

$$\boldsymbol{\beta}_0 = \boldsymbol{\beta} \quad \text{and} \quad \sigma_0 = \sigma\sqrt{\frac{k}{\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\big)}},$$

and substitution yields

$$\begin{aligned}
&\kappa(\rho, \sigma^2, \boldsymbol{\beta} \mid \rho_0) \\
&= \frac{n}{2}\left\{ k\log\left(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0)\right) - \log\left(|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0|\right) - k\log(k) \right\} \\
&= \frac{nk}{2}\log\left\{ \frac{1 + (k-2)\rho - (k-1)\rho_0\rho}{(1 + (k-1)\rho)(1-\rho)} \right\} - \frac{n}{2}\log\left\{ \frac{(1 + (k-1)\rho_0)(1-\rho_0)^{k-1}}{(1 + (k-1)\rho)(1-\rho)^{k-1}} \right\} \\
&= \kappa(\rho \mid \rho_0).
\end{aligned}$$

Therefore, the intrinsic statistic is given by

$$\delta(\rho, \rho_0) = \delta(\rho, \sigma^2, \boldsymbol{\beta}, \rho_0) = \min\big\{\kappa(\rho_0 \mid \rho), \kappa(\rho \mid \rho_0)\big\}.$$

It can be easily shown that $\kappa(\rho \mid \rho_0) \geq \kappa(\rho_0 \mid \rho)$ if and only if $\rho \in \left(-\frac{1}{k-1}, \rho_0\right]$. This completed the proof of Theorem 3.

Appendix C

The Metropolis-Hastings Algorithm from Chapter 4

Since $\boldsymbol{\gamma}$ does not follow a standard distribution, the Metropolis-Hastings algorithm can be employed. As Chib and Greenberg [21] suggested, the commonly used multivariate normal distribution is chosen as the proposal distribution, which is $\mathrm{N_q}(\mathbf{m}^{(k+1)}, \sigma_\gamma^2 V^{(k+1)})$ with

$$\mathbf{m}^{(k+1)} = \arg\max \log p(\boldsymbol{\gamma} \mid \mathbf{v}^{(k+1)}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}),$$
$$V^{(k+1)} = \{(-H)^{-1}\}_{\boldsymbol{\gamma}=\mathbf{m}}, \quad H = \frac{\partial^2 p(\boldsymbol{\gamma}|\mathbf{v}^{(k+1)}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T},$$

where $\mathbf{H}$ is the Hessian matrix and σ_γ^2 is chosen such that the average acceptance rate is between 0.25 and 0.45 (Gelman et al. [35]). For the $(k+1)$th iteration, sample $\boldsymbol{\gamma}^{(k+1)}$ by the following two steps:

Step 1: Generate a new candidate $\boldsymbol{\gamma}^{\star}$ from the proposal distribution $\mathrm{N_q}(\mathbf{m}^{(k+1)}, \sigma_\gamma^2 V^{(k+1)})$.

Step 2: Let

$$\boldsymbol{\gamma}^{(k+1)} = \begin{cases} \boldsymbol{\gamma}^{\star} & \text{if Unif}(0,1) \le \omega(\boldsymbol{\gamma}^{\star}, \boldsymbol{\gamma}^{(k)}) \\ \boldsymbol{\gamma}^{(k)} & \text{otherwise} \end{cases},$$

where $\omega(\boldsymbol{\gamma}^{\star}, \boldsymbol{\gamma}^{(k)})$ is the acceptance ratio:

$$\omega(\boldsymbol{\gamma}^{\star}, \boldsymbol{\gamma}^{(k)}) = \min\left\{1,\ \frac{p(\boldsymbol{\gamma}^{\star}|\mathbf{v}^{(k+1)}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t})}{p(\boldsymbol{\gamma}^{(k)}|\mathbf{v}^{(k+1)}, \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t})}\right\}.$$

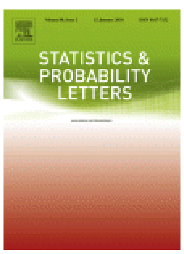
Appendix D

Letters of Permission

D.1 Copyright documentation of Chapter 2

Copyright Clearance Center RightsLink®

Home | Account Info | Help



Title: Objective Bayesian inference for the intraclass correlation coefficient in linear models

Author: Duo Zhang,Min Wang

Publication: Statistics & Probability Letters

Publisher: Elsevier

Date: June 2018

Logged in as:
DUO ZHANG
Account #:
3001445803
LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: https://www.elsevier.com/about/our-business/policies/copyright#Author-rights

BACK | CLOSE WINDOW

D.2 Copyright documentation of Chapter 3

Copyright Clearance Center RightsLink®

Home | Account Info | Help

Taylor & Francis
Taylor & Francis Group

Title: Objective Bayesian hypothesis testing and estimation for the intraclass model

Author: Duo Zhang, , , et al

Publication: Statistical Theory and Related Fields

Publisher: Taylor & Francis

Date: Jan 2, 2018

Rights managed by Taylor & Francis

Logged in as:
DUO ZHANG
Account #:
3001445803

LOGOUT

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

BACK | CLOSE WINDOW