OBJECTIVE BAYESIAN INFERENCE FOR STRESS-STRENGTH MODELS AND BAYESIAN ANOVA

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

XIAOYI MIN

Dr. Dongchu Sun, Dissertation Supervisor

DECEMBER 2012

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

OBJECTIVE BAYESIAN INFERENCE FOR
STRESS-STRENGTH MODELS AND BAYESIAN ANOVA

presented by Xiaoyi Min,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Dongchu Sun

Dr. Paul Speckman

Dr. Sounak Chakraborty

Dr. Zhuoqiong He

Dr. Jeffrey Rouder

Dedicated to my parents,

Suhong Min and Yihua Han,

and to my fiancée

Luxi Guo

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude to my advisor, Dr. Dongchu Sun. Without his guidance, inspiration, support, encouragement, and patience, this project would have never been made possible.

I would also like to thank my committee members, Drs. Paul Speckman, Sounak Chakraborty, Zhuoqiong He, and Jeffrey Rouder for their suggestions and comments on my work. A special thanks to Dr. Min Yang for pointing out the reference on orthogonal arrays.

In addition, I want to thank other faculty members in the Statistics Department who taught me, helped me, and worked with me, and I want to thank Judy, Tracy, and Kathleen for their help.

I am grateful to all the people who supported me, especially my fellow graduate students, Junlong Li, Ye Liang, Chang Xu, Yajun Liu, Haiying Wang, Na Hu, Dan Zheng, Yue Qi, Zhenyu Wang, Chester Schmaltz, Kyu Ha Lee, Aaron Berstein, Trevor Oswald, Chris Hassett, Aaron Porter, and Dan Gladish for all their help and support. Most importantly, I am indebted to my parents and Luxi for their love, understanding, and moral support through my study. It is to them that I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

OBJECTIVE BAYESIAN INFERENCE FOR STRESS-STRENGTH MODELS AND BAYESIAN ANOVA

Xiaoyi Min

Dr. Dongchu Sun, Dissertation Supervisor

ABSTRACT

First of all, for estimating the reliabilities in Weibull stress-strength models, some matching priors are derived based on a modified profile likelihood. Simulation studies show that these matching priors perform well with small sample sizes. Next, a generalized Zellner's g-prior is proposed for model selection in linear models with grouped covariates. The marginal likelihood function and a simple closed form of it are derived. The issue of computing the Bayes factors is addressed, and the performance of the Bayes factors is examined by numerical studies. Finally, the Bayes factors under the proposed prior in some 2-way ANOVA models are proved to be consistent.

Keywords: ANOVA models, Bayes factor, consistency, marginal likelihood, matching priors, modified profile likelihood, stress-strength model, Weibull distribution, Zellner's g-prior.

Chapter 1

Introduction

This dissertation consists of two distinct research projects. One is on the Bayesian inference of the reliabilities in Weibull stress-strength models in Chapter 2. The other one is on the Bayesian model selection in ANOVA models in Chapters 3 and 4.

In Chapter 2, Bayesian inference of a generalized Weibull stress-strength model (SSM) with more than one strength component is considered. For this problem, properly assigning priors for the reliabilities is challenging due to the presence of nuisance parameters. Matching priors, which are the priors matching the posterior probabilities of certain regions with their frequentist coverage probabilities, are commonly used but difficult to derive in this problem. Instead, we apply an alternative method and derive a matching prior based on a modification of the profile likelihood. Simulation studies show that this proposed prior performs well in terms of frequentist coverage and estimation even when the sample sizes are minimal. The prior is applied to two real datasets.

In Chapter 3, model selection for normal linear regression models with grouped covariates is considered under a class of Zellner's (1986) g-priors. The marginal likelihood function is derived under the proposed priors, and a simplified closed form expression is given assuming the commutativity of the projection matrices from the design matrices. As illustration, the marginal likelihood functions of the balanced q-way ANOVA models, either solely with main effects or with all interaction effects, are calculated using the closed form expression. The commutativity condition is discussed using the tool of orthogonal arrays. The approach for computing the corresponding Bayes factors is given. The performance of the proposed priors in model comparison problems is demonstrated by simulation studies on two-way ANOVA models and by two real data studies.

In Chapter 4, the prior proposed in Chapter 3 is applied to 2-way ANOVA models without interaction effects, for which the hyper-parameters in the prior need to be chosen carefully in order to yield consistent Bayes factors. Laplace approximations to the Bayes factors are given for two scenarios defined according to the asymptotic behaviors of the parameter dimensions. The Bayes factors are then proved to be consistent for both scenarios with proper choice of the hyper-parameters.

In Chapter 5, some future work about Bayesian ANOVA models is discussed.

Chapter 2

A Matching Prior Based on the Modified Profile Likelihood in a Generalized Weibull Stress-Strength Model

2.1 Introduction

In reliability studies, a stress-strength model(SSM) is often used to analyze a system that fails whenever the applied stress is greater than the strength. Denote the stress and the strength of a system by X_1 and X_2, respectively. The reliability,

$$\omega_1 = P(X_1 < X_2),$$

is often of interest to researchers. The quantity ω_1 also has interpretations under various other contexts. For example, in receiver operating characteristic (ROC) analysis, if X_1 and X_2 are, respectively, the test values of a negative instance (noise) and a positive instance (signal) chosen randomly, then the area under the ROC curve (AUC) is ω_1 (Kotz et al. 2003, Ch. 7). In a competing risks model with two risk components, suppose X_1 and X_2 are the time to failure due to the two risks, respectively. Then ω_1 is the probability that the failure is due to the first risk. Statistical inference of ω_1 is widely applicable in many areas such as engineering, clinical trials, and quality control. Various approaches on the inference of ω_1 are available, including both parametric or non-parametric, frequentist and Bayesian methods. See Basu (1985), Johnson (1988), Ghosh & Sun (1998), and Kotz et al. (2003) for reviews on many results about ω_1.

It is commonly assumed that both stress and strength follow Weibull distributions with the same shape parameter β. This assumption is convenient because ω_1 has a closed form expression. Sun et al. (1998) suggested that choosing a common shape parameter for Weibull random variables is equivalent to assuming a common scale for their logarithms. For this Weibull SSM, Sun et al. (1998) proposed several objective or noninformative priors, Kundu & Gupta (2006) summarized both frequentist and Bayesian approaches, and Krishnamoorthy & Lin (2010) proposed an interval estimation procedure based on a generalized variable approach. An exponential SSM is a special case when $\beta = 1$ and can be dated from Enis & Geisser (1971), for which Thompson & Basu (1993) derived the reference prior of ω_1.

In addition to the common SSM with one stress and one strength, the reliabilities of more complex systems have also been studied. See, for example, Kotz et al. (2003, Ch. 6.1) for a review. In this chapter, we consider the reliability of a generalized SSM, consisting of a serial system with one stress and multiple strengths,

$$\omega_1 = P(X_1 < \min(X_2, \ldots, X_I)),$$

where X_1 is the stress and $(X_2, \ldots, X_I)$ are independent strength components.

For $i = 1, \ldots, I$, let $F_i(x \mid \boldsymbol{\phi}_i)$ be the cumulative density function of X_i with parameters $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}_i$. Then ω_1 can be evaluated as a function of $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_I)$ as the following.

$$\omega_1 = \int \prod_{i=2}^{I} \left[1 - F_i(t \mid \boldsymbol{\phi}_i)\right] dF_1(t \mid \boldsymbol{\phi}_1).$$

In particular, we consider a generalized Weibull SSM by assuming that X_i satisfies the $Weibull(\eta_i, \beta)$ distribution with the density

$$f_i(x \mid \eta_i, \beta) = \frac{\beta x^{\beta-1}}{\eta_i^{\beta}} \exp(-x^{\beta}/\eta_i^{\beta}), \quad \beta, \eta_i \in (0, +\infty).$$

The distribution function is $F_i(x \mid \eta_i, \beta) = 1 - \exp[-(x/\eta_i)^{\beta}]$, and

$$\omega_1 = \frac{\eta_1^{-\beta}}{\sum_{i=1}^{I} \eta_i^{-\beta}}. \tag{2.1}$$

For this model, Hanagal (2003) studied the asymptotic distribution of the maximum likelihood estimator (MLE) of ω_1, whereas we consider its objective Bayesian inference.

The first step of a standard Bayesian procedure is to obtain the likelihood function $L(\boldsymbol{\phi} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I)$, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{in_i})$ is a random sample from $F_i(x \mid \boldsymbol{\phi}_i)$ $(i = 1, \ldots, I)$. By using an appropriate prior $\pi(\boldsymbol{\phi})$, we can calculate the posterior $\pi(\boldsymbol{\phi} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I) \propto L(\boldsymbol{\phi})\pi(\boldsymbol{\phi})$. In order to make inference on ω_1, we need to use a suitable reparametrization $\boldsymbol{\phi} \to \boldsymbol{\theta} = (\omega_1, \boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a nuisance parameter. Then we can derive the posterior for $(\omega_1, \boldsymbol{\lambda})$ by transformation rules. Finally, statistical inference about ω_1 is based on the marginal posterior of ω_1, given by

$$\pi(\omega_1 \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I) = \int \pi(\omega_1, \boldsymbol{\lambda} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I) d\boldsymbol{\lambda}.$$

Note that the posterior for $(\omega_1, \boldsymbol{\lambda})$ can also be obtained by considering the likelihood and prior under the new parametrization directly.

One difficulty of Bayesian analysis arises from eliciting the prior. Since this model has multiple parameters, a popular approach under the objective Bayesian framework is to use reference priors (Bernardo 1979, Berger & Bernardo 1992, Berger et al. 2009). Other popular priors include matching priors (Reid et al. 2003) from matching the posterior probabilities of one-sided credible intervals with their frequentist coverage probabilities, up to a certain order. In many contexts, reference priors are often first order matching priors. Unfortunately, for SSMs, matching priors may not be easy to obtain, and reference priors are usually not matching priors. See Ghosh & Sun (1998) for a review on the prior choices for several SSMs.

Furthermore, finding the marginal posterior of ω_1 requires an integration over the nuisance parameter, which often does not have a closed form expression. Numerical integrations and approximation techniques could be used but are difficult to use in general. See, for example, Ventura & Racugno (2011) for a discussion on the Laplace approximation of the marginal posterior distribution and corresponding tail area probabilities.

For models with nuisance parameters including the generalized Weibull SSM that we consider in this chapter, Ventura et al. (2009) proposed to use an appropriate pseudo-likelihood such as different modifications of the profile likelihood to eliminate the nuisance parameter $\boldsymbol{\lambda}$. Only then are matching priors on ω_1 incorporated to obtain the posterior on ω_1. This method avoids the elicitation of priors for the entire parameter and the integration on $\boldsymbol{\lambda}$. Moreover, second order matching priors are given in their paper. Hence, this approach could provide higher order matching for SSMs compared to the existing priors reviewed by Ghosh & Sun (1998). Following this path, Ventura & Racugno (2011) derived matching priors for the exponential and the normal SSMs. In this chapter, we extend the application to the generalized Weibull SSM. We should point out that this model is quite challenging to employ because there are neither simple sufficient statistics nor analytic forms for the MLE.

At the same time, the Weibull SSM is more applicable as the Weibull distribution is very popular in engineering applications, especially in the reliability context.

This chapter is organized as follows. In Section 2.2, the method proposed by Ventura et al. (2009) to handle models with nuisance parameters is briefly introduced. In Section 2.3, this method is applied to a generalized Weibull SSM, and the result for the common Weibull SSM is presented as a special case. In Section 2.4, numerical studies are carried out to demonstrate the performance of the proposed matching prior. In Section 2.5, two real datasets are analyzed. Some comments are given in Section 2.6.

2.2 Modified Profile Likelihood and the Matching Prior

To cope with the nuisance parameter $\boldsymbol{\lambda}$ in a model, several different methods have been proposed. The simplest solution is to replace the nuisance parameter $\boldsymbol{\lambda}$ by its MLE $\hat{\boldsymbol{\lambda}}$ or the restricted MLE $\hat{\boldsymbol{\lambda}}_{\omega_1}$ in the likelihood. The latter is usually referred to as the profile likelihood. Alternatively, different pseudo-likelihoods, including various modifications of the profile likelihood, can be used. Finally, Bayesian procedures, as mentioned in the introduction, can be applied to general models with nuisance parameters. The method proposed by Ventura et al. (2009) is a combination of frequentist and Bayesian approaches. Certain pseudo-likelihoods are used instead of the full likelihood as in the frequentist approach. A prior is then applied to the pseudo-likelihood, as in the Bayesian approach, to achieve the matching property. In this chapter, we only introduce two modified profile likelihoods, namely $L_{mp}(\omega_1)$ of Barndorff-Nielsen (1983) and $\bar{L}_M(\omega_1)$ by Severini (1998).

The calculation of $L_{mp}(\omega_1)$ relies on sample space derivatives. In particular, suppose that the sufficient statistic for a model can be written as $(\hat{\omega}_1, \hat{\boldsymbol{\lambda}}, \boldsymbol{a})$, where

$(\hat{\omega}_1, \hat{\boldsymbol{\lambda}})$ is the MLE of $(\omega_1, \boldsymbol{\lambda})$ and $\boldsymbol{a}$ is an ancillary statistic. The log-likelihood function $l(\omega_1, \boldsymbol{\lambda})$ can be written as $l(\omega_1, \boldsymbol{\lambda} \mid \hat{\omega}_1, \hat{\boldsymbol{\lambda}}, \boldsymbol{a})$, indicating that it is a function of $(\hat{\omega}_1, \hat{\boldsymbol{\lambda}}, \boldsymbol{a})$. Then we can calculate derivatives of the log-likelihood with respect to $(\hat{\omega}_1, \hat{\boldsymbol{\lambda}})$, which are called sample space derivatives. The sample space derivative we are interested in would be

$$l_{\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}}(\omega_1, \boldsymbol{\lambda}) = \frac{\partial^2 l(\omega_1, \boldsymbol{\lambda} \mid \hat{\omega}_1, \hat{\boldsymbol{\lambda}}, \boldsymbol{a})}{\partial \boldsymbol{\lambda} \partial \hat{\boldsymbol{\lambda}}}. \tag{2.2}$$

A detailed description of sample space derivatives can be found in Severini (2000, Ch. 6).

Based on (2.2), $L_{mp}(\omega_1)$ is defined as

$$L_{mp}(\omega_1) = L_p(\omega_1) \frac{|\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|^{\frac{1}{2}}}{|l_{\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|},$$

where $L_p(\omega_1) = L(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})$ is the profile likelihood of ω_1, and $\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) = -\partial^2 l(\omega_1, \boldsymbol{\lambda})/\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T$ is the observed Fisher information for $\boldsymbol{\lambda}$.

Write the Fisher information matrix of $(\omega_1, \boldsymbol{\lambda})$ as

$$\boldsymbol{i}(\omega_1, \boldsymbol{\lambda}) = \begin{pmatrix} \boldsymbol{i}_{\omega_1\omega_1}(\omega_1, \boldsymbol{\lambda}) & \boldsymbol{i}_{\omega_1\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) \\ \boldsymbol{i}_{\boldsymbol{\lambda}\omega_1}(\omega_1, \boldsymbol{\lambda}) & \boldsymbol{i}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) \end{pmatrix}.$$

The matching prior for ω_1 associated with $L_{mp}(\omega_1)$ given in Ventura et al. (2009) has the following form,

$$\pi(\omega_1) \propto \sqrt{\boldsymbol{i}_{\omega_1\omega_1.\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})},$$

where

$$\boldsymbol{i}_{\omega_1\omega_1.\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) = \boldsymbol{i}_{\omega_1\omega_1}(\omega_1, \boldsymbol{\lambda}) - \boldsymbol{i}_{\omega_1\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) \boldsymbol{i}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda})^{-1} \boldsymbol{i}_{\boldsymbol{\lambda}\omega_1}(\omega_1, \boldsymbol{\lambda}).$$

The corresponding posterior is

$$\pi(\omega_1 \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I) \propto L_p(\omega_1) \frac{|\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|^{\frac{1}{2}}}{|l_{\boldsymbol{\lambda};\hat{\boldsymbol{\lambda}}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|} \sqrt{\boldsymbol{i}_{\omega_1\omega_1.\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})}. \tag{2.3}$$

The calculation of $L_{mp}(\omega_1)$ might not be straightforward in many cases. For example, for the Weibull SSM, there is no analytic form for the MLE, which makes the sample space derivatives hard to obtain. For such cases, Ventura et al. (2009) suggested that their method is still applicable with other modifications to the profile likelihood. This is because all the available adjustments are equivalent to the second order, and they all reduce the score bias to $O(n^{-1})$. See Severini (2000, Ch. 9), Barndorff-Nielsen & Cox (1994, Ch. 8), and Pace & Salvan (2006) for a review. We focus on the modification obtained by Severini (1998),

$$\bar{L}_M(\omega_1) = L_p(\omega_1) \frac{|\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|^{\frac{1}{2}}}{|\boldsymbol{I}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1}; \hat{\boldsymbol{\theta}})|},$$

where $\hat{\boldsymbol{\theta}} = (\hat{\omega}_1, \hat{\boldsymbol{\lambda}})$,

$$\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0) = E_{\boldsymbol{\theta}^0}(l_{\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) l_{\boldsymbol{\lambda}}(\omega_1^0, \boldsymbol{\lambda}^0)^T),$$

with $\boldsymbol{\theta}^0 = (\omega_1^0, \boldsymbol{\lambda}^0)$ and $l_{\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) = \partial l(\omega_1, \boldsymbol{\lambda})/\partial \boldsymbol{\lambda}$. The corresponding posterior is

$$\pi^*(\omega_1 \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_I) \propto L_p(\omega_1) \frac{|\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})|^{\frac{1}{2}}}{|\boldsymbol{I}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1}; \hat{\boldsymbol{\theta}})|} \sqrt{\boldsymbol{i}_{\omega_1\omega_1.\boldsymbol{\lambda}}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1})}. \tag{2.4}$$

Applications of (2.3) and (2.4) have been illustrated in Ventura et al. (2009) and in Ventura & Racugno (2011).

2.3 A Generalized Weibull SSM

2.3.1 $\bar{L}_M(\omega_1)$ for the generalized weibull SSM

Recall that if $X_i \overset{indep}{\sim} Weibull(\eta_i, \beta)$, the expression of ω_1 is given in (2.1). We now consider the following reparametrization: for $i = 1, \ldots, I-1$, $\omega_i = \eta_i^{-\beta}/(\sum_{k=1}^{I} \eta_k^{-\beta})$, $\xi = \eta_1^{-\beta} + \cdots + \eta_I^{-\beta}$, and β stays the same. In this case, the nuisance parameter is $\boldsymbol{\lambda} = (\omega_2, \ldots, \omega_{I-1}, \xi, \beta)$. For simplicity in notation, we also let $\omega_I = 1 - \sum_{i=1}^{I-1} \omega_i = \eta_I^{-\beta}/(\sum_{k=1}^{I} \eta_k^{-\beta})$. Suppose for $i = 1, \ldots, I$, X_{ij} $(j = 1, \ldots, n_i)$ are *iid* random observations from $Weibull(\eta_i, \beta)$ distribution. Then the likelihood under the new parametrization is

$$L(\omega_1, \ldots, \omega_{I-1}, \xi, \beta) = \beta^{\sum_{i=1}^{I} n_i} \prod_{i=1}^{I} (\omega_i \xi)^{n_i} \exp\left(- \sum_{i=1}^{I} \omega_i \xi \sum_{j=1}^{n_i} X_{ij}^{\beta} \right) \prod_{i=1}^{I} \prod_{j=1}^{n_i} X_{ij}^{\beta-1}. \quad (2.5)$$

Next, we derive $\bar{L}_M(\omega_1)$ for the generalized Weibull SSM. For a related discussion on likelihood inference in Weibull models and Weibull regression models, see Ferrari et al. (2007) and Ferreira da Silva et al. (2008).

The profile likelihood L_p can be obtained by simply plugging in $\hat{\boldsymbol{\lambda}}_{\omega_1}$ in (2.5). The elements of the observed Fisher information matrix are given as follows. For $i, j = 1, \ldots, I-1$, $i \neq j$,

$$\begin{aligned}
\boldsymbol{j}_{\omega_i\omega_i} &= \frac{n_i}{\omega_i^2} + \frac{n_I}{\omega_I^2},\ \boldsymbol{j}_{\omega_i\omega_j} = \frac{n_I}{\omega_I^2},\ \boldsymbol{j}_{\omega_i\xi} = \sum_{j=1}^{n_i} X_{ij}^{\beta} - \sum_{j=1}^{n_I} X_{Ij}^{\beta},\ \boldsymbol{j}_{\xi\xi} = \frac{1}{\xi^2} \sum_{i=1}^{I} n_i, \\
\boldsymbol{j}_{\omega_i\beta} &= \xi\left(\sum_{j=1}^{n_i} X_{ij}^{\beta} \log X_{ij} - \sum_{j=1}^{n_I} X_{Ij}^{\beta} \log X_{Ij} \right),\ \boldsymbol{j}_{\xi\beta} = \sum_{i=1}^{I} \omega_i \sum_{j=1}^{n_i} X_{ij}^{\beta} \log X_{ij}, \\
\boldsymbol{j}_{\beta\beta} &= \frac{1}{\beta^2} \sum_{i=1}^{I} n_i + \xi \sum_{i=1}^{I} \omega_i \sum_{j=1}^{n_i} X_{ij}^{\beta} (\log X_{ij})^2.
\end{aligned}$$

The explicit form of $|\boldsymbol{j}_{\boldsymbol{\lambda\lambda}}(\omega_1, \boldsymbol{\lambda})|$ is quite complicated, and thus not calculated in this chapter. The following lemma gives $\boldsymbol{I}(\omega_1, \hat{\boldsymbol{\lambda}}_{\omega_1}; \hat{\boldsymbol{\theta}})$.

Lemma 1. *Let γ denote Euler's constant, and let $\Gamma(x)$, $\psi(x)$, and $\psi_1(x)$ be the Gamma, digamma, and trigamma functions, respectively. The elements of the matrix $\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0)$ are given as follows. For $i, j = 2, \ldots, I-1$, $i \neq j$,*

$$
\begin{aligned}
\boldsymbol{I}_{\omega_i\omega_i} &= \frac{\beta\xi}{\beta^0\left(\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(1+\frac{\beta}{\beta^0}\Big)\left[\frac{n_i}{\left(\omega_i^0\right)^{1+\frac{\beta}{\beta^0}}}+\frac{n_I}{\left(\omega_I^0\right)^{1+\frac{\beta}{\beta^0}}}\right],\\
\boldsymbol{I}_{\omega_i\omega_j} &= \frac{n_I\beta\xi}{\beta^0\left(\xi^0\right)^{\frac{\beta}{\beta^0}}\left(\omega_I^0\right)^{1+\frac{\beta}{\beta^0}}}\Gamma\Big(1+\frac{\beta}{\beta^0}\Big),\\
\boldsymbol{I}_{\omega_i\xi} &= \frac{\beta\xi}{\beta^0\left(\xi^0\right)^{1+\frac{\beta}{\beta^0}}}\Gamma\Big(1+\frac{\beta}{\beta^0}\Big)\left[\frac{n_i}{\left(\omega_i^0\right)^{\frac{\beta}{\beta^0}}}-\frac{n_I}{\left(\omega_I^0\right)^{\frac{\beta}{\beta^0}}}\right],\\
\boldsymbol{I}_{\xi\omega_i} &= \frac{\beta}{\beta^0\left(\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(1+\frac{\beta}{\beta^0}\Big)\left[\frac{n_i\omega_i}{\left(\omega_i^0\right)^{1+\frac{\beta}{\beta^0}}}-\frac{n_I\omega_I}{\left(\omega_I^0\right)^{1+\frac{\beta}{\beta^0}}}\right],\\
\boldsymbol{I}_{\omega_i\beta} &= \frac{n_i\beta\xi}{\left(\beta^0\right)^2\left(\omega_i^0\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\left[\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_i^0\xi^0\right)\right]\\
&- \frac{n_I\beta\xi}{\left(\beta^0\right)^2\left(\omega_I^0\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\left[\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_I^0\xi^0\right)\right],\\
\boldsymbol{I}_{\beta\omega_i} &= \frac{n_i\beta\omega_i\xi}{\left(\beta^0\right)^2\left(\omega_i^0\right)^{1+\frac{\beta}{\beta^0}}\left(\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\left[\frac{\beta^0}{\beta}+\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_i^0\xi^0\right)\right]-\frac{n_i}{\beta^0\omega_i^0}\\
&- \frac{n_I\beta\omega_I\xi}{\left(\beta^0\right)^2\left(\omega_I^0\right)^{1+\frac{\beta}{\beta^0}}\left(\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\left[\frac{\beta^0}{\beta}+\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_I^0\xi^0\right)\right]+\frac{n_I}{\beta^0\omega_I^0},\\
\boldsymbol{I}_{\xi\xi} &= \frac{\beta}{\beta^0}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\sum_{i=1}^{I}\frac{n_i\omega_i}{\left(\omega_i^0\right)^{\frac{\beta}{\beta^0}}\left(\xi^0\right)^{1+\frac{\beta}{\beta^0}}},\\
\boldsymbol{I}_{\xi\beta} &= \frac{\beta}{\left(\beta^0\right)^2}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\sum_{i=1}^{I}\frac{n_i\omega_i}{\left(\omega_i^0\xi^0\right)^{\frac{\beta}{\beta^0}}}\left[\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_i^0\xi^0\right)\right],\\
\boldsymbol{I}_{\beta\xi} &= \frac{\beta}{\left(\beta^0\right)^2}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\sum_{i=1}^{I}\frac{n_i\omega_i\xi}{\left(\omega_i^0\right)^{\frac{\beta}{\beta^0}}\left(\xi^0\right)^{1+\frac{\beta}{\beta^0}}}\left[\frac{\beta^0}{\beta}+\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\left(\omega_i^0\xi^0\right)\right]\\
&- \sum_{i=1}^{I}\frac{n_i}{\xi^0\beta^0},\\
\boldsymbol{I}_{\beta\beta} &= \frac{1}{\left(\beta^0\right)^2}\sum_{i=1}^{I}n_i\big[\gamma+\log\left(\omega_i^0\xi^0\right)\big]+\sum_{i=1}^{I}\frac{n_i\beta\omega_i\xi}{\left(\beta^0\right)^3\left(\omega_i^0\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(\frac{\beta}{\beta^0}+1\Big)\\
&\quad\left\{\psi_1\Big(\frac{\beta}{\beta^0}+1\Big)+\left[\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\omega_i^0\xi^0\right]\left[\frac{\beta^0}{\beta}+\psi\Big(\frac{\beta}{\beta^0}+1\Big)-\log\omega_i^0\xi^0\right]\right\}.
\end{aligned}
$$

Proof. To calculate $|\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0)|$, the partial derivatives of the log-likelihood are:

$$\begin{aligned}
\frac{\partial l}{\partial \omega_i} &= \frac{n_i}{\omega_i} - \frac{n_I}{\omega_I} - \xi \sum_{j=1}^{n_i} X_{ij}^{\beta} + \xi \sum_{j=1}^{n_I} X_{Ij}^{\beta}, \text{ for } i = 1, \ldots, I-1, \\
\frac{\partial l}{\partial \xi} &= \frac{\sum_{i=1}^{I} n_i}{\xi} - \sum_{i=1}^{I} \omega_i \sum_{j=1}^{n_i} X_{ij}^{\beta}, \\
\frac{\partial l}{\partial \beta} &= \frac{\sum_{i=1}^{I} n_i}{\beta} + \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log X_{ij} - \xi \sum_{i=1}^{I} \omega_i \sum_{j=1}^{n_i} X_{ij}^{\beta} \log X_{ij}.
\end{aligned}$$

We notice that $E_{\boldsymbol{\theta}^0}[l_{\boldsymbol{\lambda}}(\omega_1^0, \boldsymbol{\lambda}^0)] = \mathbf{0}$. Therefore,

$$\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0) = E_{\boldsymbol{\theta}^0}\left[l_{\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) l_{\boldsymbol{\lambda}}(\omega_1^0, \boldsymbol{\lambda}^0)^T\right] - E_{\boldsymbol{\theta}^0}\left[l_{\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda})\right] E_{\boldsymbol{\theta}^0}\left[l_{\boldsymbol{\lambda}}(\omega_1^0, \boldsymbol{\lambda}^0)^T\right]. \quad (2.6)$$

Next, we derive $I_{\omega_i\omega_i}$ (for $i = 1, \ldots, I-1$) to illustrate our calculation. Since X_{ij}'s ($j = 1, \ldots, n_i$) and X_{Ij}'s ($j = 1, \ldots, n_I$) are mutually independent, most of the terms in the product cancel out due to (2.6). Hence,

$$\begin{aligned}
I_{\omega_i\omega_i} &= \xi\xi^0 \Bigg\{ \sum_{j=1}^{n_i} \left[E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta+\beta^0}\right) - E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta}\right) E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta^0}\right) \right] \\
&+ \sum_{j=1}^{n_I} \left[E_{\boldsymbol{\theta}^0}\left(X_{Ij}^{\beta+\beta^0}\right) - E_{\boldsymbol{\theta}^0}\left(X_{Ij}^{\beta}\right) E_{\boldsymbol{\theta}^0}\left(X_{Ij}^{\beta^0}\right) \right] \Bigg\}.
\end{aligned}$$

Suppose the true parameters are $\boldsymbol{\theta}^0$, $\omega_i^0 \xi^0 X_{ij}^{\beta^0} \sim \textit{Exponential}(1)$, so

$$E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta+\beta^0}\right) - E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta}\right) E_{\boldsymbol{\theta}^0}\left(X_{ij}^{\beta^0}\right) = \frac{1}{\left(\omega_i^0 \xi^0\right)^{1+\frac{\beta}{\beta^0}}} \Big(E\big(Z^{1+\frac{\beta}{\beta^0}}\big) - E(Z) E\big(Z^{\frac{\beta}{\beta^0}}\big) \Big),$$

where Z is an $\textit{Exponential}(1)$ random variable. Thus,

$$\begin{aligned}
E_{\boldsymbol{\theta}^0}(X_{ij}^{\beta+\beta^0}) - E_{\boldsymbol{\theta}^0}(X_{ij}^{\beta}) E_{\boldsymbol{\theta}^0}(X_{ij}^{\beta^0}) &= \frac{1}{\left(\omega_i^0 \xi^0\right)^{1+\frac{\beta}{\beta^0}}} \left[\Gamma\Big(2 + \frac{\beta}{\beta^0}\Big) - \Gamma\Big(1 + \frac{\beta}{\beta^0}\Big) \right] \\
&= \frac{\beta}{\beta^0 \left(\omega_i^0 \xi^0\right)^{1+\frac{\beta}{\beta^0}}} \Gamma\Big(1 + \frac{\beta}{\beta^0}\Big),
\end{aligned}$$

which leads to

$$I_{\omega_i\omega_i} = \frac{\beta\xi}{\beta^0\left(\xi^0\right)^{\frac{\beta}{\beta^0}}}\Gamma\Big(1+\frac{\beta}{\beta^0}\Big)\left[\frac{n_i}{\left(\omega_i^0\right)^{1+\frac{\beta}{\beta^0}}} + \frac{n_I}{\left(\omega_I^0\right)^{1+\frac{\beta}{\beta^0}}}\right].$$

The other elements can also be calculated using the same approach. □

The expression of $|\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0)|$ is rather complicated and thus omitted. Finally, $\bar{L}_M(\omega_1)$ for this model can be obtained by combining the results above.

2.3.2 The matching prior

To find the matching prior corresponding to $\bar{L}_M(\omega_1)$, we first calculate the components of the Fisher information matrix as follows. For $i, j = 1, \ldots, I-1$, $i \neq j$,

$$\begin{aligned}
\boldsymbol{i}_{\omega_i\omega_i} &= \frac{n_i}{\omega_i^2} + \frac{n_I}{\omega_I^2},\ \boldsymbol{i}_{\omega_i\omega_j} = \frac{n_I}{\omega_I^2},\ \boldsymbol{i}_{\omega_i\xi} = \frac{n_i}{\omega_i\xi} - \frac{n_I}{\omega_I\xi}, \\
\boldsymbol{i}_{\omega_i\beta} &= \frac{n_i(1-\gamma-\log\omega_i\xi)}{\omega_i\beta} - \frac{n_I(1-\gamma-\log\omega_I\xi)}{\omega_I\beta},\ \boldsymbol{i}_{\xi\xi} = \frac{1}{\xi^2}\sum_{i=1}^{I} n_i \\
\boldsymbol{i}_{\xi\beta} &= \frac{1}{\beta\xi}\sum_{i=1}^{I} n_i(1-\gamma-\log\omega_i\xi),\ \boldsymbol{i}_{\beta\beta} = \frac{1}{\beta^2}\sum_{i=1}^{I} n_i\left[\frac{\pi^2}{6} + (1-\gamma-\log\omega_i\xi)^2\right].
\end{aligned}$$

We obtain

$$\boldsymbol{i}_{\omega_1\omega_1,\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) = \frac{\frac{\pi^2}{6}(\sum_{i=1}^{I} n_i)\frac{n_1}{\omega_1^2}}{\frac{\pi^2}{6}(\sum_{i=1}^{I} n_i)[n_1\sum_{i=2}^{I}\frac{\omega_i^2}{n_i} + (1-\omega_1)^2] + n_1(\sum_{i=2}^{I}\omega_i\log\frac{\omega_1}{\omega_i})^2}.$$

Hence,

$$\pi(\omega_1) \propto \frac{1}{\omega_1}\left\{\frac{\pi^2}{6}\sum_{i=1}^{I} n_i\Big[n_1\sum_{i=2}^{I}\frac{\hat{\omega}_i^2(\omega_1)}{n_i} + (1-\omega_1)^2\Big] + n_1\Big(\sum_{i=2}^{I}\hat{\omega}_i(\omega_1)\log\frac{\omega_1}{\hat{\omega}_i(\omega_1)}\Big)^2\right\}^{-\frac{1}{2}}, \quad (2.7)$$

where $\hat{\omega}_i(\omega_1)$ denotes that restricted MLE of ω_i given ω_1. Combining $\pi(\omega_1)$ and $\bar{L}_M(\omega_1)$, the posterior (2.4) of ω_1 can be obtained. Note that this prior depends on the restricted MLE, and thus on the data, which might seem unappealing as the data is used twice. However, this is common for matching priors. As shown by Wasserman (2000) and reviewed in Reid et al. (2003), data-dependent priors are sometimes necessary for the posterior to achieve good frequentist properties, especially in terms of coverage probabilities.

2.3.3 The Weibull SSM with one strength component

The previous results can be applied to the common Weibull SSM with only one strength component. Following similar notation and parametrization, the likelihood of (ω_1, ξ, β) is

$$
\begin{aligned}
L(\omega_1, \xi, \beta) \; &= \; \beta^{n_1+n_2} \left(\prod_{i=1}^{2} \prod_{j=1}^{n_i} X_{ij} \right)^{\beta-1} (\omega_1 \xi)^{n_1} [(1-\omega_1)\xi]^{n_2} \\
& \quad \exp\left[-\omega_1 \xi \sum_{j=1}^{n_1} X_{1j}^{\beta} - (1-\omega_1)\xi \sum_{j=1}^{n_2} X_{2j}^{\beta} \right]. \qquad (2.8)
\end{aligned}
$$

The $(\boldsymbol{\lambda}, \boldsymbol{\lambda})$-block of the observed Fisher information is

$$
\begin{aligned}
& \boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda}) \\
= & \begin{pmatrix} \frac{n_1+n_2}{\xi^2} & \omega_1 \sum_{j=1}^{n_1} X_{1j}^{\beta} \log X_{1j} + (1-\omega_1) \sum_{j=1}^{n_2} X_{2j}^{\beta} \log X_{2j} \\ * & \frac{n_1+n_2}{\beta^2} + \omega_1 \xi \sum_{j=1}^{n_1} X_{1j}^{\beta} (\log X_{1j})^2 + (1-\omega_1)\xi \sum_{j=1}^{n_2} X_{2j}^{\beta} (\log X_{2j})^2 \end{pmatrix}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
|\boldsymbol{j}_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\omega_1, \boldsymbol{\lambda})| \; &= \; \frac{(n_1+n_2)^2}{\xi^2 \beta^2} - \left[\omega_1 \sum_{j=1}^{n_1} X_{1j}^{\beta} \log X_{1j} + (1-\omega_1) \sum_{j=1}^{n_2} X_{2j}^{\beta} \log X_{2j} \right]^2 \\
& + \; \frac{n_1+n_2}{\xi} \left[\omega_1 \sum_{j=1}^{n_1} X_{1j}^{\beta} (\log X_{1j})^2 + (1-\omega_1) \sum_{j=1}^{n_2} X_{2j}^{\beta} (\log X_{2j})^2 \right], \quad (2.9)
\end{aligned}
$$

and

$$
\begin{aligned}
|\boldsymbol{I}(\omega_1, \boldsymbol{\lambda}; \boldsymbol{\theta}^0)| &= \frac{\beta^2 \xi}{(\beta^0)^4 \xi^0} \big[\Gamma(\frac{\beta}{\beta^0}+1)\big]^2 \Bigg\{ \frac{n_1 n_2 \omega_1 (1-\omega_1)}{\big[\omega_1^0 (1-\omega_1^0)(\xi^0)^2\big]^{\frac{\beta}{\beta^0}}} \Big(\log \frac{\omega_1^0}{1-\omega_1^0} \Big)^2 \\
&+ \ \psi_1(\frac{\beta}{\beta^0}+1) \Big\{ \frac{n_1 \omega_1}{(\omega_1^0 \xi^0)^{\frac{\beta}{\beta^0}}} + \frac{n_2(1-\omega_1)}{\big[(1-\omega_1^0)\xi^0\big]^{\frac{\beta}{\beta^0}}} \Big\}^2 \Bigg\} \\
&+ \ \frac{\beta}{(\beta^0)^3 \xi^0} \Gamma(\frac{\beta}{\beta^0}+1) \Bigg\{ (n_1+n_2)\big[\gamma + \psi(\frac{\beta}{\beta^0}+1)\big] \Big\{ \frac{n_1 \omega_1}{(\omega_1^0 \xi^0)^{\frac{\beta}{\beta^0}}} + \frac{n_2(1-\omega_1)}{\big[(1-\omega_1^0)\xi^0\big]^{\frac{\beta}{\beta^0}}} \Big\} \\
&+ \ n_1 n_2 \log \Big(\frac{\omega_1^0}{1-\omega_1^0} \Big) \Big\{ \frac{1-\omega_1}{\big[(1-\omega_1^0)\xi^0\big]^{\frac{\beta}{\beta^0}}} - \frac{\omega_1}{(\omega_1^0 \xi^0)^{\frac{\beta}{\beta^0}}} \Big\} \Bigg\}. \qquad (2.10)
\end{aligned}
$$

The matching prior can be derived as a special case of (2.7). Let $a = n_1/(n_1+n_2)$,

$$
\pi(\omega_1) \propto \frac{1}{\omega_1(1-\omega_1)} \left[\frac{\pi^2}{6} + a(1-a) \left(\log \frac{\omega_1}{1-\omega_1} \right)^2 \right]^{-\frac{1}{2}}. \qquad (2.11)
$$

Surprisingly, unlike the general case, the matching prior for this model is independent of the data. It is the same as the marginal of the reference prior for the orders $\{\omega_1, (\xi, \beta)\}$ and $\{\omega_1, \beta, \xi\}$ derived by Sun et al. (1998). Interestingly, the reference prior is not even a first-order matching prior for the original likelihood, whereas its marginal is a second-order matching prior for $\bar{L}_M(\omega_1)$. Finally, the posterior is (2.4), where the related terms are given in (2.8), (2.9), (2.10), and (2.11).

The posterior propriety is rather difficult to prove in general because the MLE and the restricted MLE do not have closed form expressions. However, according to Severini (2007), $\bar{L}_M(\omega_1)$ is an approximation to the integrated likelihood function with respect to any conditional prior $\pi(\boldsymbol{\lambda} \mid \omega_1)$. Therefore, the posterior we propose for the common Weibull SSM is an approximation to the marginal posterior based on the reference prior derived by Sun et al. (1998), which was proved to be proper when $n_1 + n_2 \geq 3$.

For the generalized SSM, first note that

$$\left\{\frac{\pi^2}{6}\sum_{i=1}^{I} n_i\Big[n_1\sum_{i=2}^{I}\frac{\hat{\omega}_i^2(\omega_1)}{n_i}+(1-\omega_1)^2\Big]+n_1\Big(\sum_{i=2}^{I}\hat{\omega}_i(\omega_1)\log\frac{\omega_1}{\hat{\omega}_i(\omega_1)}\Big)^2\right\}^{-\frac{1}{2}}\leq\frac{C}{1-\omega_1}.$$

Moreover, $\bar{L}_M(\omega_1)\omega_1^{-1}(1-\omega_1)^{-1}$ is an approximation to the marginal posterior with prior $\pi(\omega_1,\ldots,\omega_{I-1},\xi,\beta)=\beta^{-I}\xi^{-1}\omega_1^{-1}(1-\omega_1)^{-1}$, which is proper when $n_1+\cdots+n_I\geq I+1$.

2.4 Numerical Studies

We carry out a simulation study to examine the performance of our prior. Note that the posterior that we derive needs to be evaluated numerically because it involves the MLE $\hat{\boldsymbol{\theta}}=(\hat{\omega}_1,\hat{\boldsymbol{\lambda}})$ and the restricted MLE $\hat{\boldsymbol{\lambda}}_{\omega_1}$, which do not have closed form expressions. For a SSM, since the parameter of interest ω_1 is only defined on $(0,1)$, evaluating the posterior of ω_1 on a fine grid in $(0,1)$ is sufficient for posterior inference.

First, we consider the model with only one strength component. The simulation settings are the same as those in Sun et al. (1998). In particular, the parameters (η_1,η_2,β) take the values $(3,2,0.5)$, $(3,2,1.0)$, and $(3,2,1.5)$ whereas the sample sizes (n_1,n_2) take the values $(2,2)$, $(2,3)$, $(5,5)$, and $(10,10)$. Under each setting, we generate $10,000$ samples of stresses and strengths $(\boldsymbol{X}_1,\boldsymbol{X}_2)$ independently, where $\boldsymbol{X}_i=(X_{i1},\ldots,X_{in_i})$ is a sample of size n_i from $Weibull(\eta_i,\beta)$ $(i=1,2)$. For each set of data $(\boldsymbol{X}_1,\boldsymbol{X}_2)$, we find the posterior $\pi^*(\omega_1\mid\boldsymbol{X}_1,\boldsymbol{X}_2)$ and $\bar{L}_M(\omega_1)$. Note that $\bar{L}_M(\omega_1)$ can be regarded as the posterior with a uniform prior on ω_1. The corresponding posterior mode is compared with the MLE and the mode of $\bar{L}_M(\omega_1)$ in terms of bias and mean squared error (MSE). The results are summarized in Table 2.1. In Table 2.2, we summarize and compare the frequentist coverage probabilities of the one-sided 0.05 and 0.95 credible (confidence) intervals derived from four methods.

The posterior credible intervals from our prior are labeled π^*, the credible intervals based on $\bar{L}_M(\omega_1)$ are denoted with $\bar{L}_M$, and 'Frequentist' refers to the frequentist confidence intervals based on the observed Fisher information matrices. We also include the results for the matching prior π_m given by Sun et al. (1998), which we cite from their paper.

Moreover, we consider the generalized SSM with two strength components. Three sets of parameters are used for $(\eta_1, \eta_2, \eta_3, \beta)$: $(3, 2, 2, 0.5)$, $(3, 2, 2, 1.0)$, and $(3, 2, 2, 1.5)$. The sample sizes (n_1, n_2, n_3) take the values $(2, 2, 2)$, $(2, 3, 3)$, $(3, 3, 3)$, $(5, 5, 5)$, and $(10, 10, 10)$. Similarly, for each setting, we generate $10,000$ independent samples of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$, and we obtain $\pi^*(\omega_1 \mid \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ and $\bar{L}_M(\omega_1)$ for each set of data. The posterior mode is compared with the MLE and the mode of $\bar{L}_M(\omega_1)$ in Table 2.3. The 0.05 and the 0.95 one-sided posterior credible intervals based on π^* are compared with those based on $\bar{L}_M(\omega_1)$ and the frequentist confidence intervals in Table 2.4.

In terms of coverage probabilities, the Bayesian credible intervals based on our prior perform much better than the confidence intervals based on the MLE and the credible interval based on $\bar{L}_M(\omega_1)$. This is most likely due to the small sample size that we use. In both simulations, the relative frequencies of coverage of our proposed posterior are very close to the nominal levels even when the samples sizes are minimal. This observation is similar to that of Ventura & Racugno (2011) on the exponential and normal SSMs. However, for the Weibull SSM with only one strength component, π_m is slightly better than our prior in the sense that the coverage probabilities converge to the nominal levels faster as the sample sizes grow. In terms of estimation bias and MSE, the posterior mode from our method performs similarly to the MLE and the mode of $\bar{L}_M(\omega_1)$ under all settings. When the sample sizes become larger, the biases are quite stable, while the MSEs become smaller. Our proposed prior seems to perform reasonably well in terms of both point estimation and coverage probabilities of the credible regions.

2.5 Data Analysis

We first analyze the carbon fiber strength data from Badar & Priest (1982). It consists of the tensile strength (in GPa) measured for single carbon fibers and impregnated 1000-carbon fiber tows at different gauge lengths. Surles & Padgett (1998, 2001) analyzed the data with a generalized Rayleigh distribution. Kundu & Gupta (2006) reanalyzed a subset of the data, and estimated $\omega_1 = P(X_1 < X_2)$, where X_1 and X_2 are the strengths of single fibers of 20mm and 10mm, respectively. The subsample sizes are $n_1 = 69$ and $n_2 = 63$. After subtracting 0.75 from both samples, they performed the Kolmogorov-Smirnov (K-S) test and found that the assumption of Weibull distributions with a common shape parameter is quite suitable.

We use the Weibull SSM and apply our method. The posterior mode of ω_1 is 0.7637 and the credible interval of ω_1 is $(0.6859, 0.8223)$. They are very similar to the MLE and confidence intervals obtained in Kundu & Gupta (2006). The value of ω_1 is significantly greater than 0.5, which suggests that 20mm fibers have greater strength than 10mm fibers, as expected.

We also analyze the silicon carbide fiber strength data provided in Kovacs (1996) using the Weibull SSM with two strength components. It includes the tensile strengths of fibers tested at different gauge lengths. We estimate $\omega_1 = P(X_1 < \min(X_2, X_3))$, where X_1, X_2, and X_3 are the strength of fibers at 50, 10, and 5mm, respectively. The corresponding sample sizes are $n_1 = 32$, $n_2 = 45$, and $n_3 = 31$. The data set is presented below.

50mm Fibers: 0.935, 1.087, 1.558, 1.721, 1.766, 1.836, 1.926, 2.134, 2.278, 2.287, 2.293, 2.332, 2.419, 2.461, 2.507, 2.538, 2.542, 2.615, 2.686, 2.715, 2.780, 2.818, 2.838, 2.842, 2.849, 2.905, 2.963, 2.968, 3.395, 3.438, 3.686, 3.773.

10mm Fibers: 1.608, 1.687, 2.192, 2.194, 2.386, 2.665, 2.805, 2.970, 3.068, 3.153, 3.272, 3.315, 3.324, 3.449, 3.605, 3.623, 3.652, 3.677, 3.704, 3.745, 3.774, 3.783, 3.882, 3.943, 3.963, 4.041, 4.152, 4.238, 4.275, 4.286, 4.313, 4.331, 4.615, 4.620, 4.677, 4.791, 4.893, 5.061, 5.081, 5.116, 5.119, 5.175, 5.206, 5.378, 5.482.

5mm Fibers: 1.655, 2.681, 2.965, 2.958, 3.146, 3.542, 3.713, 3.830, 3.867, 3.874, 3.890, 3.958, 4.047, 4.074, 4.091, 4.098, 4.170, 4.256, 4.288, 4.384, 4.573, 4.687, 4.974, 5.076, 5.130, 5.153, 5.191, 5.430, 5.723, 5.948, 6.239.

Following a procedure similar to Kundu & Gupta (2006), we first fit the Weibull models to the three sets of data separately. Table 2.5 summarizes the estimated shape and scale parameters, log-likelihood values, K-S statistics, and the corresponding p-values. Results from the K-S tests show that the Weibull models fit the data quite well. The estimated shape parameters are quite close and K-S statistics suggest that the Weibull models with a common shape parameter fit the data well. The results based on the Weibull models with a common shape parameter are given in Table 2.6. Note that the log-likelihood values with or without this assumption are very close. The MLE of ω_1 is 0.8169. The posterior mode based on our method is 0.8185 and the 95% credible interval is $(0.7179, 0.8784)$. This again suggests that the longer fibers in this study have smaller strengths with a high probability.

2.6 Comments

In this chapter, we consider the problem of evaluating the reliability in a generalized Weibull SSM with more than one strength component. This model contains a large number of nuisance parameters, so traditional Bayesian inference is quite challenging. Therefore, we apply the method proposed in Ventura et al. (2009) and Ventura & Racugno (2011) using the modified profile likelihood and the corresponding matching priors.

The simulation results illustrate that the prior performs well in terms of both point and interval estimations, even for very small sample sizes. Instead of multidimensional numerical integration or Markov chain Monte Carlo, this method relies on relatively easy optimization steps. By replacing $L_{mp}(\omega_1)$ by $\bar{L}_M(\omega_1)$ as suggested in Ventura & Racugno (2011), the calculation of sample space derivatives is avoided, which makes this method applicable to many complicated statistical problems.

Table 2.1: Estimation bias and MSE for posterior mode(π^*), MLE, and mode of $\bar{L}_M$ in the Weibull SSM.

		π^*		MLE		$\bar{L}_M$	
(η_1, η_2, β)	(n_1, n_2)	Bias	MSE	Bias	MSE	Bias	MSE
$(3, 2, 0.5)$	$(2, 2)$	-0.004	0.11	0.001	0.095	0.008	0.068
	$(2, 3)$	0.008	0.092	0.020	0.079	0.017	0.059
	$(5, 5)$	-0.008	0.037	-0.005	0.034	-0.001	0.029
	$(10, 10)$	-0.004	0.016	-0.002	0.015	0	0.014
$(3, 2, 1.0)$	$(2, 2)$	-0.012	0.106	-0.005	0.092	0.012	0.066
	$(2, 3)$	0.005	0.091	0.021	0.078	0.025	0.060
	$(5, 5)$	-0.010	0.035	-0.005	0.032	0.003	0.027
	$(10, 10)$	-0.008	0.015	-0.005	0.014	-0.001	0.013
$(3, 2, 1.5)$	$(2, 2)$	-0.012	0.096	-0.002	0.083	0.022	0.060
	$(2, 3)$	-0.001	0.084	0.018	0.073	0.028	0.056
	$(5, 5)$	-0.019	0.032	-0.011	0.029	0.001	0.025
	$(10, 10)$	-0.010	0.013	-0.006	0.013	0	0.012

Table 2.2: Frequentist coverage probabilities of 0.05 and 0.95 credible (confidence) intervals of ω_1 for the Weibull SSM. (Monte Carlo SE=0.0022)

		π^*		π_m		$Frequentist$		$\bar{L}_M$	
(η_1, η_2, β)	(n_1, n_2)	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95
$(3, 2, 0.5)$	$(2, 2)$	0.0553	0.9612	0.012	0.997	0.1964	0.7541	0.0088	1
	$(2, 3)$	0.0553	0.9444	0.029	0.973	0.1903	0.8061	0.0232	0.9885
	$(5, 5)$	0.0556	0.9416	0.042	0.959	0.0998	0.8613	0.0487	0.9595
	$(10, 10)$	0.0536	0.9436	0.050	0.953	0.0708	0.9030	0.0513	0.9543
$(3, 2, 1.0)$	$(2, 2)$	0.0584	0.9698	0.014	0.999	0.1711	0.7252	0.0313	1
	$(2, 3)$	0.0629	0.9489	0.029	0.974	0.1722	0.7826	0.0411	0.9978
	$(5, 5)$	0.0597	0.9465	0.042	0.959	0.0910	0.8489	0.0579	0.9686
	$(10, 10)$	0.0544	0.9491	0.049	0.951	0.0632	0.8934	0.0552	0.9617
$(3, 2, 1.5)$	$(2, 2)$	0.0551	0.9887	0.015	0.999	0.1443	0.7120	0.0432	1
	$(2, 3)$	0.0631	0.9550	0.029	0.975	0.1467	0.7593	0.0541	1
	$(5, 5)$	0.0585	0.9441	0.043	0.960	0.0749	0.8282	0.0616	0.9776
	$(10, 10)$	0.0528	0.9498	0.048	0.953	0.0535	0.8847	0.0564	0.9653

Table 2.3: Estimation bias and MSE for posterior mode(π^*), MLE, and mode of $\bar{L}_M$ in the Weibull SSM with 2 strength components.

		π^*		MLE		$\bar{L}_M$	
$(\eta_1, \eta_2, \eta_3, \beta)$	(n_1, n_2, n_3)	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$(3, 2, 2, 0.5)$	$(2, 2, 2)$	-0.004	0.071	-0.001	0.071	0.040	0.050
	$(2, 3, 3)$	0.007	0.061	0.025	0.059	0.047	0.045
	$(3, 3, 3)$	-0.008	0.042	-0.006	0.041	0.026	0.033
	$(5, 5, 5)$	-0.005	0.020	-0.004	0.020	0.018	0.018
	$(10, 10, 10)$	-0.002	0.008	-0.002	0.008	0.010	0.008
$(3, 2, 2, 1.0)$	$(2, 2, 2)$	-0.004	0.063	-0.002	0.062	0.046	0.046
	$(2, 3, 3)$	0.005	0.056	0.022	0.055	0.049	0.043
	$(3, 3, 3)$	-0.011	0.036	-0.010	0.035	0.027	0.029
	$(5, 5, 5)$	-0.007	0.017	-0.007	0.017	0.017	0.015
	$(10, 10, 10)$	-0.004	0.007	-0.004	0.007	0.009	0.007
$(3, 2, 2, 1.5)$	$(2, 2, 2)$	-0.010	0.053	-0.009	0.052	0.045	0.041
	$(2, 3, 3)$	-0.001	0.045	0.017	0.045	0.048	0.037
	$(3, 3, 3)$	-0.015	0.030	-0.014	0.029	0.026	0.026
	$(5, 5, 5)$	-0.011	0.014	-0.010	0.014	0.015	0.013
	$(10, 10, 10)$	-0.006	0.006	-0.006	0.005	0.008	0.005

Table 2.4: Frequentist coverage probabilities of 0.05 and 0.95 credible(confidence) intervals of ω_1 in the Weibull SSM with 2 strength components. (Monte Carlo SE=0.0022)

		π^*		$Frequentist$		$\bar{L}_M$	
$(\eta_1, \eta_2, \eta_3, \beta)$	(n_1, n_2, n_3)	0.05	0.95	0.05	0.95	0.05	0.95
$(3, 2, 2, 0.5)$	$(2, 2, 2)$	0.0490	0.9691	0.1232	0.6624	0.0618	1
	$(2, 3, 3)$	0.0492	0.9444	0.1254	0.7507	0.0656	0.9903
	$(3, 3, 3)$	0.0513	0.9400	0.0859	0.7436	0.0687	0.9850
	$(5, 5, 5)$	0.0546	0.9395	0.0611	0.8200	0.0679	0.9696
	$(10, 10, 10)$	0.0500	0.9488	0.0453	0.8837	0.0647	0.9652
$(3, 2, 2, 1.0)$	$(2, 2, 2)$	0.0548	0.9802	0.1061	0.6380	0.0778	1
	$(2, 3, 3)$	0.0570	0.9442	0.1067	0.7233	0.0818	0.9973
	$(3, 3, 3)$	0.0549	0.9436	0.0718	0.7220	0.0798	0.9906
	$(5, 5, 5)$	0.0537	0.9459	0.0501	0.8059	0.0771	0.9750
	$(10, 10, 10)$	0.0546	0.9457	0.0424	0.8671	0.0721	0.9660
$(3, 2, 2, 1.5)$	$(2, 2, 2)$	0.0528	0.9937	0.0831	0.6047	0.0901	1
	$(2, 3, 3)$	0.0511	0.9478	0.0849	0.7148	0.0858	0.9995
	$(3, 3, 3)$	0.0562	0.9461	0.0581	0.6969	0.0895	0.9961
	$(5, 5, 5)$	0.0535	0.9434	0.0405	0.7811	0.0795	0.9789
	$(10, 10, 10)$	0.0496	0.9495	0.0332	0.8597	0.0684	0.9724

Table 2.5: Parameter estimates, log-likelihood, K-S statistics, and p-values of the fitted weibull models to the silicon carbide fibers data.

Data Set	Shape	Scale	Log-Likelihood	K-S Statistic	p-value
50mm	4.414	2.738	-31.305	0.1149	0.7498
10mm	4.703	4.241	-61.753	0.0772	0.9327
5mm	4.895	4.625	-43.307	0.1142	0.7715

Table 2.6: Parameter estimates, log-likelihood, K-S statistics, and p-values of the fitted weibull models (equal shape parameters) to the silicon carbide fibers data.

Data Set	Shape	Scale	Log-Likelihood	K-S Statistic	p-value
50mm	4.664	2.752	-31.386	0.1162	0.737
10mm	4.664	4.238	-61.756	0.0763	0.9382
5mm	4.664	4.606	-43.366	0.1192	0.7267

Chapter 3

Bayesian Model Selection for ANOVA Models

3.1 Introduction

Model comparison for linear models is a common problem in statistical inference, one that can be applied to many areas such as biology, psychology, chemistry, and economics. Under the frequentist framework, the problem of model comparison involves two distinct approaches that are dependent on the number of models being compared. When there are two models under comparison, the approach of hypothesis testing is applied, using, for example, the p-value, whereas when the comparison involves more than two models, quite different tools of model selection, such as AIC and C_p, are used. In contrast, the Bayesian approach is conceptually the same regardless of the number of models being compared. Berger & Pericchi (2001) discussed the advantages of the Bayesian approach over the classical frequentist methods in model comparison problems. Various procedures exist for Bayesian model comparison. A common approach is to use the Bayes factor based on posterior model probabilities (Kass & Raftery 1995). Suppose that we are comparing r models M_i $(i = 1, \ldots, r)$ for the data $\boldsymbol{y}$, with the density of $\boldsymbol{y}$ under M_i being $f_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is the unknown model parameter. Suppose that in M_i $(i = 1, \ldots, r)$, the prior distribution for $\boldsymbol{\theta}_i$ is $\pi_i(\boldsymbol{\theta}_i)$. Define the marginal likelihood for model M_i,

$$m_i(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}_i} f_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i. \tag{3.1}$$

The Bayes factor of M_i to M_j $(i, j = 1, \ldots, r)$ is defined as

$$B_{ij} = \frac{m_i(\boldsymbol{y})}{m_j(\boldsymbol{y})}. \tag{3.2}$$

In this study, we consider the Bayes factor for a general class of linear mixed models.

An important issue in using the Bayes factor approach is the choice of prior distributions for the regression coefficients and the variances. A large amount of literature already addresses this problem. In general, Berger & Pericchi (2001) suggested that proper vague priors should be avoided since they will yield undesirable results, and using improper priors requires care as they may produce an indeterminate Bayes factor. A typical approach for linear models is the conventional prior approach (Berger & Pericchi 2001) introduced by Jeffreys (1961), which was further developed by Zellner & Siow (1980) and Zellner (1986). The Zellner-Siow (1980) prior, which assigns a multivariate Cauchy prior to the coefficients, has been extensively discussed. Zellner's (1986) g-prior is another widely used prior, which gives a closed form expression for the marginal likelihood given a hyper-parameter g. Recent work by Bayarri & García-Donato (2007) extended the Zellner-Siow prior to deal with general linear models. Liang et al. (2008) reviewed the choices of g for Zellner's g-prior, and they compared fixed g-priors with mixtures of g-priors. Other priors that have been proposed include intrinsic priors (Berger & Pericchi 1996) and expected posterior priors (Pérez & Berger 2002). In this study, a modification of Zellner's g-prior is proposed, and the performance of the corresponding Bayes factors is studied.

There are two motivations for this study. The first one arises from the difficulties of Zellner's g-prior when it is applied to the linear regression models with the following setting:

$$\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_m\boldsymbol{\beta}_m + \boldsymbol{\epsilon}, \tag{3.3}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector, for $j = 0, 1, \ldots, m$, $\boldsymbol{X}_j$ is an $n \times p_j$ known design matrix of full column rank, $\boldsymbol{\beta}_j$ is a $p_j \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\epsilon} \sim N_n\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n\big)$. This setting is common for linear models with grouped covariates, and in particular it is useful for ANOVA models.

Two difficulties exist when using Zellner's g-prior. First, Zellner's g-prior requires the design matrix to have full column rank, which is often not true in ANOVA models.

Example 1. *Consider a two-way ANOVA model with the main effects of factors A and B and their interaction, where A has p_1 levels, B has p_2 levels, and each combination of levels has k replicates. Let*

$$\boldsymbol{y} \;\; = \;\; \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{\epsilon}, \tag{3.4}$$

where $\boldsymbol{y}$ is an $n = p_1p_2k$ dimensional vector, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ are vectors of unknown effects with dimensions p_1, p_2, and p_1p_2 respectively, $\boldsymbol{X}_1 = \boldsymbol{I}_{p_1} \otimes \boldsymbol{1}_{p_2} \otimes \boldsymbol{1}_k$, $\boldsymbol{X}_2 = \boldsymbol{1}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{1}_k$, and $\boldsymbol{X}_3 = \boldsymbol{I}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{1}_k$, and $\boldsymbol{\epsilon} \sim N_n\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n\big)$. Clearly, the design matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ is not of full column rank. Zellner's g-prior on $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$ cannot be computed since it involves the inversion of matrix $\boldsymbol{X}'\boldsymbol{X}$. Of course, one could consider a reparametrization, but the prior might not have a simple form, and the interpretation of the parameters might be difficult.

Secondly, Zellner's g-prior for a model in (3.3) lacks flexibility since one hyper-parameter g controls the priors for several different groups of the parameters $\boldsymbol{\beta}_j$'s

$(j = 1, \ldots, m)$. Thus the variances of the priors for different $\boldsymbol{\beta}_j$'s are limited to changing at the same scale as g changes. In this study, we try to bring more flexibility to the priors by using more hyper-parameters g_j $(j = 1, \ldots, m)$, so that g_j controls the prior on $\boldsymbol{\beta}_j$ independently.

In traditional linear model theory, the regression coefficients or effects can be classified into two categories: fixed effects and random effects. If the effects are unknown constants, they are called fixed effects. For example, in the linear regression model (3.3), the regression coefficients $\boldsymbol{\beta}_j$ $(j = 1, \ldots, m)$ are commonly considered as fixed effects. Random effects are those considered to be random variables. For example, the effects in ANOVA models such as $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ in (3.4) are often considered to be random effects. Moreover, if fixed effects and random effects both exist in a linear model (3.3), the model is called a linear mixed model. In many real life applications, however, it is not easy to decide whether an effect is fixed or random. Thus, it is appealing to develop a unified treatment to model comparison problems for fixed effects and random effects.

To date, for Bayesian model comparison, the literature has primarily addressed the properties of the prior distribution for fixed effect models. Very few discussed models with random effects, among which García-Donato & Sun (2007) considered the Bayes factor for testing a one-way random effect model under both intrinsic priors and divergence based priors. Sun et al. (2012) is perhaps the first to consider the unification of priors for fixed effects and random effects. The second motivation of this study is to show that Bayes factors with suitable priors can accommodate the model comparison problem for all cases: fixed effect models, random effect models, and mixed models.

This chapter is organized as follows. In Section 3.2, the marginal likelihood of model (3.3) is derived under the proposed prior, and a simpler closed form expression is calculated under the commutativity condition. In Section 3.3, the result from

Section 3.2 is demonstrated by the application to two special cases. In Section 3.4, the commutativity condition in Section 3.2 is discussed. In Section 3.5, the scheme for computing the Bayes factors under the proposed prior is given. In Section 3.6, the performance of the proposed prior is demonstrated using simulation studies. In Section 3.7, the proposed method is applied to two real data studies. Finally, we conclude with stating our findings.

3.2 Main Results

3.2.1 The Proposed Priors

Consider the model of (3.3). We propose a modification of Zellner's g-prior to this model and derive the marginal likelihood $m(\boldsymbol{y})$ under the proposed priors so that the Bayes factor can be calculated to compare the models of interest. Specifically, if all the $\boldsymbol{\beta}_j$'s are fixed effects, we choose the right Haar prior for the parameters $(\boldsymbol{\beta}_0, \sigma^2)$,

$$\pi(\boldsymbol{\beta}_0, \sigma^2) = \frac{1}{\sigma^2}. \tag{3.5}$$

As we have discussed earlier, we consider the independent multivariate Cauchy priors for the conditional prior of $\boldsymbol{\beta}_j$ given $(\boldsymbol{\beta}_0, \sigma^2)$,

$$[\boldsymbol{\beta}_j \mid \boldsymbol{\beta}_0, \sigma^2] = \frac{\Gamma(\frac{p_j+1}{2})|X_j'X_j|^{\frac{1}{2}}}{\pi^{\frac{p_j+1}{2}}(n\sigma^2)^{\frac{p_j}{2}}} \left(1 + \frac{1}{n\sigma^2}\boldsymbol{\beta}_j' X_j' X_j \boldsymbol{\beta}_j\right)^{-\frac{p_j+1}{2}}, \tag{3.6}$$

namely, $Cauchy_{p_j}(\mathbf{0}, (\boldsymbol{X}_j'\boldsymbol{X}_j)/n\sigma^2)$. This prior can be decomposed as the hierarchical structure,

$$\boldsymbol{\beta}_j \mid \sigma^2, g_j^* \overset{indep}{\sim} N_{p_j}\big(\mathbf{0}, n g_j^* \sigma^2 (\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\big), \tag{3.7}$$

$$g_j^* \overset{indep}{\sim} \text{Inv-Gamma}(1/2, 1/2), \tag{3.8}$$

for $j = 1, \ldots, m$. Substituting ng_j^* with g_j, the hierarchical structure is equivalent to

$$\begin{aligned}
\boldsymbol{\beta}_j \mid \sigma^2, g_j &\overset{indep}{\sim} N_{p_j}\big(\mathbf{0}, g_j\sigma^2(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\big), &(3.9)\\
g_j &\overset{indep}{\sim} \text{Inv-Gamma}(1/2, n/2), &(3.10)
\end{aligned}$$

for $j = 1, \ldots, m$. Of course, for a regression model, we could just use the first stage of this hierarchical structure (3.9) for fixed g_j as the prior for $\boldsymbol{\beta}_j$. This is the Zellner's g-prior. At the same time, if some of the effects in $\boldsymbol{\beta}_j$, $j = 1, \ldots, m$, are unobserved random effects, then the corresponding priors of $\boldsymbol{\beta}_j$ in the first stage of this hierarchical structure (3.9) are often a part of the model, with unknown parameter g_j, the ratio of the corresponding variance component and error variance σ^2. Different prior choices for the second stage of the hierarchical structure are just priors for g_j. Other hyper-priors on g_j could also be used as the second stage of the hierarchical structure (3.10) (Liang et al. 2008, Maruyama & George 2011, Guo & Speckman 2009).

3.2.2 The Effective Sample Size

One benefit of using Bayes factor is that the marginal likelihood automatically penalizes for model complexity, which is often called the Bayesian "Ockham's razor effect" (Jefferys & Berger 1992). This is because integrating the likelihood on a parameter space with higher dimensions will result in a more diffuse predictive distribution for the data. In our problem, the scale of the hyper-prior on g_j in (3.10) determines the scale of this penalty. Thus, choosing an appropriate scale in (3.10) is crucial for obtaining good Bayes factors that penalize complex models properly. In particular, Berger et al. (2010) suggested to use 'The Effective Sample Size' (TESS) instead of n in the context of model selection, which motivates us to discuss the properties of the Bayes factors if we choose Inv-Gamma($1/2$, nb_j) distribution as the prior on g_j (for $j = 1, \ldots, m$).

Here, we introduce 'TESS' for linear models in Berger et al. (2010) and give an example. They consider linear models of the form

$$\boldsymbol{Y} = \boldsymbol{X}^*\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.11}$$

where $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ known, with dimensions $\boldsymbol{Y}_{n\times 1}$, $\boldsymbol{X}^*_{n\times q}$, $\boldsymbol{\alpha}_{q\times 1}$, $\boldsymbol{X}_{n\times p}$, $\boldsymbol{\beta}_{p\times 1}$, $\boldsymbol{\epsilon}_{n\times 1}$ and $\boldsymbol{\Sigma}_{n\times n}$. They assume that $\boldsymbol{X}^*\boldsymbol{\alpha}$ is the common part in every model, but $\boldsymbol{X}\boldsymbol{\beta}$ differ from model to model. It is also assumed that $\boldsymbol{\alpha}$ is transformed so that it is orthogonal to $\boldsymbol{\beta}$ in every model. They then propose TESS for the parameter $\boldsymbol{\xi} = \boldsymbol{v}\boldsymbol{\beta}$, where $\boldsymbol{v}$ is a $1 \times p$ vector, is

$$n^e = \frac{|\boldsymbol{v}|^2}{\boldsymbol{v}\boldsymbol{F}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{F}\boldsymbol{v}'}, \tag{3.12}$$

where $\boldsymbol{F}$ is the $p \times p$ diagonal matrix with entries

$$f_{ii} = \max_j\{|X_{ji}/\sigma_j|\}, \tag{3.13}$$

and $\boldsymbol{\Sigma} = \boldsymbol{\sigma}\boldsymbol{S}\boldsymbol{\sigma}$, with $\boldsymbol{\sigma}$ being the diagonal matrix of standard deviations σ_j and $\boldsymbol{S}$ being the correlation matrix. The next example illustrates TESS in the two-way ANOVA model that we consider earlier.

Example 1 (Continued). *For this model, μ is the common parameter. As the design matrix $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ does not have full column rank, we cannot apply (3.12) directly. Thus, we derive TESS for the three parts of the design matrix separately. For example, for each element in $\boldsymbol{\beta}_1$, it is easy to derive that TESS is p_2k. This also makes intuitive sense in that there are p_2k replicates under each level of the corresponding factor. Similarly, we can derive that TESS is p_1k and k for each element in $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, respectively. We will see that replacing n by TESS in (3.10) for this model is a reasonable choice and results in Bayes factors with good properties.*

3.2.3 Marginal Likelihood

Next, we derive the marginal likelihood under our proposed prior. For the convenience of calculation, we use the following notation:

$$\boldsymbol{\beta} = \big(\boldsymbol{\beta}_0', \boldsymbol{\beta}^{*'}\big)', \ \boldsymbol{\beta}^* = \big(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m'\big)', \ \boldsymbol{g} = \big(g_1, g_2, \ldots, g_m\big), \tag{3.14}$$

$$\boldsymbol{P}_j = \boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j', \quad j = 0, 1, \ldots, m, \tag{3.15}$$

$$\boldsymbol{X} = \big(\boldsymbol{X}_0, \boldsymbol{X}^*\big), \ \boldsymbol{X}^* = \big(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m\big), \tag{3.16}$$

$$\boldsymbol{M} = diag\big(\mathbf{0}_{p_0 \times p_0}, \boldsymbol{M}_1\big), \tag{3.17}$$

$$\boldsymbol{M}_1 = diag\Big(\frac{1}{g_1}\boldsymbol{X}_1'\boldsymbol{X}_1, \frac{1}{g_2}\boldsymbol{X}_2'\boldsymbol{X}_2, \ldots, \frac{1}{g_m}\boldsymbol{X}_m'\boldsymbol{X}_m\Big). \tag{3.18}$$

Notice that Zellner's g-prior on $\boldsymbol{\beta}^*$ given (σ^2, g) is $N(\mathbf{0}, \sigma^2(\frac{1}{g}\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1})$, and our proposed prior (3.9) on $\boldsymbol{\beta}^*$ given $(\sigma^2, \boldsymbol{g})$ is $N(\mathbf{0}, \sigma^2\boldsymbol{M}_1^{-1})$, which has a block diagonal variance-covariance matrix. Comparing to the Zellner's g-prior, our proposed prior treats each part of the parameters $\boldsymbol{\beta}_j$ $(j = 1, \ldots, m)$ separately, so the priors of different parts have more flexibility and are independent of each other.

The likelihood function of $(\boldsymbol{\beta}, \sigma^2)$ based on $\boldsymbol{y}$ is

$$f(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big[-\frac{1}{2\sigma^2}\big(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\big)'\big(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\big)\Big]. \tag{3.19}$$

The conditional prior of $\boldsymbol{\beta}$ given $(\boldsymbol{g}, \sigma^2)$ is

$$[\boldsymbol{\beta} \mid \boldsymbol{g}, \sigma^2] = \prod_{j=1}^{m}[\boldsymbol{\beta}_j \mid g_j, \sigma^2] = \left[\prod_{j=1}^{m}\frac{\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}}{(2\pi g_j\sigma^2)^{p_j/2}}\right]\exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{\beta}'\boldsymbol{M}\boldsymbol{\beta}\Big). \tag{3.20}$$

Theorem 1. *The marginal likelihood function given* $\boldsymbol{g}$ *is*

$$m(\boldsymbol{y} \mid \boldsymbol{g}) = \frac{\Gamma\left(\frac{n-p_0}{2}\right)}{(\pi\,\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n-p_0}{2}}\big|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}\big|^{1/2}}\prod_{j=1}^{m}\frac{\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}}{g_j^{p_j/2}}, \tag{3.21}$$

where

$$\boldsymbol{R} \;=\; \boldsymbol{I}_n - \boldsymbol{X}\big(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big)^{-1}\boldsymbol{X}'. \tag{3.22}$$

(b) If the prior on $\boldsymbol{g}$ is $\pi(\boldsymbol{g})$, the marginal likelihood of $\boldsymbol{y}$ will be

$$m(\boldsymbol{y}) \;=\; \frac{\Gamma\left(\frac{n-p_0}{2}\right)}{\pi^{\frac{n-p_0}{2}}}\prod_{j=1}^{m}\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}\int\frac{\prod_{j=1}^{m} g_j^{-p_j/2}}{(\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n-p_0}{2}}\big|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big|^{1/2}}\pi(\boldsymbol{g})d\boldsymbol{g}. \tag{3.23}$$

Proof. Note that

$$\begin{aligned} m(\boldsymbol{y}\mid\boldsymbol{g}) \;&=\; \int_0^\infty \frac{1}{\sigma^2}\frac{1}{(2\pi\sigma^2)^{n/2}}\prod_{j=1}^{m}\Big[\frac{\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}}{(2\pi g_j\sigma^2)^{p_j/2}}\Big] \\ &\quad \int \exp\Big\{-\frac{1}{2\sigma^2}\Big[\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\big)'\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\big)+\boldsymbol{\beta}'\boldsymbol{M}\boldsymbol{\beta}\Big]\Big\}d\boldsymbol{\beta}d\sigma^2. \end{aligned} \tag{3.24}$$

If we write $\tilde{\boldsymbol{\beta}} = \big(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big)^{-1}\boldsymbol{X}'\boldsymbol{y}$,

$$\begin{aligned} &\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\big)'\big(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\big)+\boldsymbol{\beta}'\boldsymbol{M}\boldsymbol{\beta} \\ =\;& (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'\big(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big)(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})+\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}, \end{aligned} \tag{3.25}$$

where $\boldsymbol{R}$ is defined by (3.22). Therefore,

$$\begin{aligned} & m(\boldsymbol{y}\mid\boldsymbol{g}) \\ =\;& \int_0^\infty \frac{\exp\Big(-\frac{1}{2\sigma^2}\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}\Big)}{(2\pi)^{\frac{n-p_0}{2}}(\sigma^2)^{\frac{n-p_0}{2}+1}\big|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big|^{\frac{1}{2}}}\prod_{j=1}^{m}\Big(\frac{\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}}{g_j^{p_j/2}}\Big)d(\sigma^2) \\ =\;& \frac{\Gamma\Big(\frac{n-p_0}{2}\Big)}{(2\pi)^{\frac{n-p_0}{2}}\big|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\big|^{\frac{1}{2}}}\prod_{j=1}^{m}\Big(\frac{\big|\boldsymbol{X}_j'\boldsymbol{X}_j\big|^{1/2}}{g_j^{p_j/2}}\Big)\left(\frac{2}{\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}}\right)^{\frac{n-p_0}{2}}. \end{aligned} \tag{3.26}$$

This proves (3.21). □

3.2.4 Commutativity Assumption

In (3.21), we need to compute both

$$\left|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\right| \text{ and } \left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}\right)^{-1}. \tag{3.27}$$

The computation could be extensive if the dimension of $\boldsymbol{M}$, $(p_0 + p_1 + \cdots + p_m)$, is large. Interestingly, under the following commutativity condition of the projection matrices,

$$\boldsymbol{P}_i\boldsymbol{P}_j = \boldsymbol{P}_j\boldsymbol{P}_i, \ \forall\ i, j, \tag{3.28}$$

there is a simple form for (3.27), so we can write a real closed form expression of the marginal likelihood without inverting the matrix in (3.27) numerically. This might be a strong condition, but it is satisfied for balanced ANOVA models.

Next, we derive this simple form and the marginal likelihood under this commutativity condition. For $\boldsymbol{\gamma} \subseteq \{0, 1, \ldots, m\}$, define

$$\begin{aligned}
\boldsymbol{P}_{\boldsymbol{\gamma}} &= \prod_{j\in\boldsymbol{\gamma}} \boldsymbol{P}_j, && (3.29)\\
\boldsymbol{A}_{\boldsymbol{\gamma}} &= \prod_{j\in\boldsymbol{\gamma}} \boldsymbol{P}_j \prod_{j'\in\{0,1,\ldots,m\}\setminus\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_{j'}), && (3.30)\\
p_{\boldsymbol{\gamma}} &= rank(\boldsymbol{A}_{\boldsymbol{\gamma}}). && (3.31)
\end{aligned}$$

Condition (3.28) guarantees that both (3.29) and (3.30) are well defined. For convenience, we define the collection of all nonempty subsets of $\{1, \ldots, m\}$ by

$$\boldsymbol{\Gamma} = \Big\{\{1\}, \{2\}, \ldots, \{m\}, \{1, 2\}, \ldots, \{m-1, m\}, \ldots, \{1, 2, \ldots, m\}\Big\}, \tag{3.32}$$

and we use $\boldsymbol{\Gamma}$ as the index set.

Lemma 2. *Suppose that (3.28) holds.*

(a) Both $\boldsymbol{P}_{\boldsymbol{\gamma}}$ *in (3.29) and* $\boldsymbol{A}_{\boldsymbol{\gamma}}$ *in (3.30) are projection matrices.*

(b) For any $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}^* \subseteq \{0, 1, \ldots, m\}$*, we have* $\boldsymbol{A}_{\boldsymbol{\gamma}}\boldsymbol{A}_{\boldsymbol{\gamma}^*} = \boldsymbol{0}$.

(c) We have the expression for the determinant,

$$\big|\boldsymbol{I}_n + \sum_{j=1}^{m} g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\big| = \prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \big(1 + \sum_{j\in\boldsymbol{\gamma}} g_j\big)^{p_{\boldsymbol{\gamma}}}. \tag{3.33}$$

(d) We have the expression for the inverse,

$$\Big[\boldsymbol{I}_n + \sum_{j=1}^{m} g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\Big]^{-1} = \boldsymbol{I}_n + \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}, \tag{3.34}$$

where

$$u_{\boldsymbol{\gamma}} = (-1)^k \sum_{(j_1, j_2, \ldots, j_k)} \Big(\frac{g_{j_1}}{1 + g_{j_1}} \frac{g_{j_2}}{1 + g_{j_1} + g_{j_2}} \cdots \frac{g_{j_k}}{1 + g_{j_1} + \cdots + g_{j_k}}\Big), \tag{3.35}$$

with $k = |\boldsymbol{\gamma}|$ *and the summation taken over all the possible permutations of* $\boldsymbol{\gamma}$.

(e) $u_{\boldsymbol{\gamma}}$ *defined in (3.35) satisfies the following property: for any* $\boldsymbol{\gamma}_0 \in \boldsymbol{\Gamma}$,

$$\sum_{\emptyset\neq\boldsymbol{\gamma}\subseteq\boldsymbol{\gamma}_0} u_{\boldsymbol{\gamma}} = -1 + \frac{1}{1 + \sum_{j\in\boldsymbol{\gamma}_0} g_j}. \tag{3.36}$$

Proof. Parts (a) and (b) are easy. For Part (c), note the following decompositions.

$$\boldsymbol{I}_n = \sum_{\boldsymbol{\gamma}\subseteq\{0,\ldots,m\}} \boldsymbol{A}_{\boldsymbol{\gamma}}, \qquad (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j = \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}: j\in\boldsymbol{\gamma}} \boldsymbol{A}_{\boldsymbol{\gamma}}. \tag{3.37}$$

Therefore,

$$\boldsymbol{I}_n + \sum_{j=1}^{m} g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j = \sum_{\boldsymbol{\gamma}\notin\boldsymbol{\Gamma}} \boldsymbol{A}_{\boldsymbol{\gamma}} + \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \boldsymbol{A}_{\boldsymbol{\gamma}}(1 + \sum_{j:j\in\boldsymbol{\gamma}} g_j). \tag{3.38}$$

$\forall \boldsymbol{\gamma} \subseteq \{0, \ldots, m\}$, $\boldsymbol{A}_{\boldsymbol{\gamma}}$ is idempotent and symmetric, whose eigenvalues are $p_{\boldsymbol{\gamma}}$ 1's and $(n - p_{\boldsymbol{\gamma}})$ 0's. Therefore, there is an $n \times p_{\boldsymbol{\gamma}}$ matrix $\boldsymbol{B}_{\boldsymbol{\gamma}}$ (if $p_{\boldsymbol{\gamma}} = 0$, we let $\boldsymbol{B}_{\boldsymbol{\gamma}}$ be a null matrix) such that $\boldsymbol{A}_{\boldsymbol{\gamma}} = \boldsymbol{B}_{\boldsymbol{\gamma}} \boldsymbol{B}'_{\boldsymbol{\gamma}}$ and $\boldsymbol{B}'_{\boldsymbol{\gamma}} \boldsymbol{B}_{\boldsymbol{\gamma}} = \boldsymbol{I}_{p_{\boldsymbol{\gamma}}}$. Note that for $\boldsymbol{\gamma}^* \neq \boldsymbol{\gamma}$, $\boldsymbol{B}'_{\boldsymbol{\gamma}} \boldsymbol{B}_{\boldsymbol{\gamma}^*} = \boldsymbol{0}_{p_{\boldsymbol{\gamma}} \times p_{\boldsymbol{\gamma}^*}}$. Further, if $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, write $\boldsymbol{C}_{\boldsymbol{\gamma}} = \sqrt{1 + \sum_{j \in \boldsymbol{\gamma}} g_j} \boldsymbol{B}_{\boldsymbol{\gamma}}$; if $\boldsymbol{\gamma} \notin \boldsymbol{\Gamma}$, define $\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{B}_{\boldsymbol{\gamma}}$. We then combine all $\boldsymbol{C}_{\boldsymbol{\gamma}}$'s side-by-side into an $n \times n$ matrix $\boldsymbol{C}$ and get

$$\left| \boldsymbol{I}_n + \sum_{j=1}^{m} g_j (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_j \right| = |\boldsymbol{C}\boldsymbol{C}'| = |\boldsymbol{C}'\boldsymbol{C}|, \tag{3.39}$$

and (3.33) follows by noting that $\boldsymbol{C}'\boldsymbol{C}$ is a block diagonal matrix with the diagonal parts being $(1 + \sum_{j \in \boldsymbol{\gamma}} g_j) \boldsymbol{I}_{p_{\boldsymbol{\gamma}}}$ if $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, and $\boldsymbol{I}_{p_{\boldsymbol{\gamma}}}$, otherwise.

For Part (d), note that

$$(\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_j = \begin{cases} (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}}, & if \;\; j \in \boldsymbol{\gamma}, \\ (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma} \cup \{j\}}, & if \;\; j \notin \boldsymbol{\gamma}. \end{cases} \tag{3.40}$$

Consider the product of $(\boldsymbol{I}_n + \sum_{j=1}^{m} g_j (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_j)$ and $(\boldsymbol{I}_n + \sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}})$ where $u_{\boldsymbol{\gamma}}$ is defined as in (3.35), we prove that the coefficient before each term $(\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}}$ is zero. For $(\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\{j\}}$, we need to show that

$$g_j + u_{\{j\}} + g_j u_{\{j\}} = 0, \tag{3.41}$$

which is true since

$$u_{\{j\}} = -\frac{g_j}{1 + g_j}. \tag{3.42}$$

For $(\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}}$ with $|\boldsymbol{\gamma}| = k \geq 2$, we need to show that

$$u_{\boldsymbol{\gamma}} + \sum_{j \in \boldsymbol{\gamma}} g_j (u_{\boldsymbol{\gamma}} + u_{\boldsymbol{\gamma} \setminus \{j\}}) = 0, \tag{3.43}$$

which is true because

$$u_{\boldsymbol{\gamma}} = -\frac{\sum_{j\in\boldsymbol{\gamma}} g_j u_{\boldsymbol{\gamma}\setminus\{j\}}}{1+\sum_{j\in\boldsymbol{\gamma}} g_j}. \tag{3.44}$$

Thus, (3.34) and (3.35) are proved.

For Part (e), without loss of generality, we only prove (3.36) for $\boldsymbol{\gamma}_0 = \{1,\ldots,k\}$. In fact,

$$\begin{aligned}
&\sum_{\emptyset\neq\boldsymbol{\gamma}\subseteq\{1,\ldots,k\}} u_{\boldsymbol{\gamma}} \\
&= \sum_{j_1}\left(\frac{-g_{j_1}}{1+g_{j_1}}\left(1+\sum_{j_2}\frac{-g_{j_2}}{1+g_{j_1}+g_{j_2}}\left(\cdots\left(1+\sum_{j_k}\frac{-g_{j_k}}{1+g_{j_1}+\cdots+g_{j_k}}\right)\cdots\right)\right.\right. \\
&= \sum_{j_1}\left(\frac{-g_{j_1}}{1+g_{j_1}}\left(\cdots\left(1+\sum_{j_{k-1}}\frac{-g_{j_{k-1}}}{1+g_{j_1}+\cdots+g_{j_{k-1}}}\left(\frac{1+g_{j_1}+\cdots+g_{j_{k-1}}}{1+g_1+\cdots+g_k}\right)\cdots\right)\right.\right. \\
&= \sum_{j_1}\left(\frac{-g_{j_1}}{1+g_{j_1}}\left(1+\sum_{j_2}\frac{-g_{j_2}}{1+g_{j_1}+g_{j_2}}\left(\cdots\left(\frac{1+g_{j_1}+\cdots+g_{j_{k-2}}}{1+g_1+\cdots+g_k}\right)\cdots\right)\right.\right..
\end{aligned}$$

By the induction, we have

$$\begin{aligned}
\sum_{\emptyset\neq\boldsymbol{\gamma}\subseteq\{1,\ldots,k\}} u_{\boldsymbol{\gamma}} &= \sum_{j_1}\left(\frac{-g_{j_1}}{1+g_{j_1}}\frac{1+g_{j_1}}{1+g_1+\cdots+g_k}\right) \\
&= \frac{-(g_1+\cdots+g_k)}{1+g_1+\cdots+g_k} = -1+\frac{1}{1+g_1+\cdots+g_k}.
\end{aligned}$$

The lemma is proved. □

Theorem 2. *Assume the commutativity condition (3.28). Then $m(\boldsymbol{y}\mid\boldsymbol{g})$, given in (3.21), has the expression,*

$$m(\boldsymbol{y}\mid\boldsymbol{g}) = \frac{\Gamma\left(\frac{n-p_0}{2}\right)}{(\pi\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n-p_0}{2}}\left|\boldsymbol{X}_0'\boldsymbol{X}_0\right|^{1/2}}\prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\frac{1}{\left(1+\sum_{j\in\boldsymbol{\gamma}} g_j\right)^{p_{\boldsymbol{\gamma}}/2}}, \tag{3.45}$$

where $\boldsymbol{R}$ defined in (3.22) has the expression,

$$\boldsymbol{R} = \boldsymbol{I}_n - \boldsymbol{P}_0 + \sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_0) \boldsymbol{P}_{\boldsymbol{\gamma}}, \tag{3.46}$$

and $u_{\boldsymbol{\gamma}}$ is defined in (3.35).

Proof. To calculate $\boldsymbol{R}$, we write

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{I}_{p_0} & \boldsymbol{0} \\ -(\boldsymbol{X}^*)' \boldsymbol{X}_0 (\boldsymbol{X}_0' \boldsymbol{X}_0)^{-1} & \boldsymbol{I}_{p_1 + \cdots + p_m} \end{pmatrix}, \tag{3.47}$$

where $\boldsymbol{X}^*$ is defined in (3.16). Then

$$\boldsymbol{R} \;=\; \boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{D}' \left[\boldsymbol{D}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}'\right]^{-1} \boldsymbol{D}\boldsymbol{X}'. \tag{3.48}$$

We can also show that

$$\left[\boldsymbol{D}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}'\right]^{-1} = diag\Big((\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1},\ \left(\boldsymbol{X}^{*'}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^* + \boldsymbol{M}_1\right)^{-1}\Big).$$

Define $\tilde{\boldsymbol{X}} = (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^*$. We get

$$\boldsymbol{R} = (\boldsymbol{I}_n - \boldsymbol{P}_0) - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}'. \tag{3.49}$$

Using the fact that for invertible matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Delta}$,

$$(\boldsymbol{\Phi} + \boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\omega}')^{-1} = \boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\boldsymbol{\omega}(\boldsymbol{\Delta}^{-1} + \boldsymbol{\omega}'\boldsymbol{\Phi}^{-1}\boldsymbol{\omega})^{-1}\boldsymbol{\omega}'\boldsymbol{\Phi}^{-1}, \tag{3.50}$$

we get

$$(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1} = \boldsymbol{M}_1^{-1} - \boldsymbol{M}_1^{-1}\tilde{\boldsymbol{X}}'(\boldsymbol{I}_n + \tilde{\boldsymbol{X}}\boldsymbol{M}_1^{-1}\tilde{\boldsymbol{X}}')^{-1}\tilde{\boldsymbol{X}}\boldsymbol{M}_1^{-1}. \tag{3.51}$$

Define $\boldsymbol{O} = \tilde{\boldsymbol{X}}\boldsymbol{M}_1^{-1}\tilde{\boldsymbol{X}}'$. Then

$$\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}' = \boldsymbol{O} - \boldsymbol{O}(\boldsymbol{I}_n + \boldsymbol{O})^{-1}\boldsymbol{O} = \boldsymbol{I}_n - (\boldsymbol{I}_n + \boldsymbol{O})^{-1}. \tag{3.52}$$

Also,

$$\begin{aligned} \boldsymbol{I}_n + \boldsymbol{O} = \boldsymbol{I}_n + (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^*\boldsymbol{M}_1^{-1}\boldsymbol{X}^{*'}(\boldsymbol{I}_n - \boldsymbol{P}_0) & \\ = \boldsymbol{I}_n + (\boldsymbol{I}_n - \boldsymbol{P}_0)\Big(\sum_{j=1}^m g_j \boldsymbol{P}_j\Big)(\boldsymbol{I}_n - \boldsymbol{P}_0) \;&=\; \boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j. \end{aligned} \tag{3.53}$$

Applying (3.53) and (3.34) to (3.52), we get

$$\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}' = -\sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}. \tag{3.54}$$

This proves (3.46). Next, we calculate $|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}|$. For $\boldsymbol{D}$ defined in (3.47),

$$|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}| = |\boldsymbol{D}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}'| \;=\; |\boldsymbol{X}_0'\boldsymbol{X}_0|\,|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1|. \tag{3.55}$$

Using the identity $|\boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\omega}' + \boldsymbol{\Phi}| = |\boldsymbol{\Delta}|\,|\boldsymbol{\Phi}|\,|\boldsymbol{\Delta}^{-1} + \boldsymbol{\omega}'\boldsymbol{\Phi}^{-1}\boldsymbol{\omega}|$, we have

$$\begin{aligned} |\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}| &= |\boldsymbol{X}_0'\boldsymbol{X}_0|\,|\boldsymbol{M}_1|\,|\boldsymbol{I}_n + \tilde{\boldsymbol{X}}\boldsymbol{M}_1^{-1}\tilde{\boldsymbol{X}}'| \\ &= \prod_{j=0}^m |\boldsymbol{X}_j'\boldsymbol{X}_j|\Big(\prod_{j=1}^m g_j^{-p_j}\Big)\Big|\boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\Big|. \end{aligned} \tag{3.56}$$

By Part (c) of Lemma 2,

$$|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}| = \prod_{j=0}^m |\boldsymbol{X}_j'\boldsymbol{X}_j|\Big(\prod_{j=1}^m g_j^{-p_j}\Big)\prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\Big(1 + \sum_{j\in\boldsymbol{\gamma}} g_j\Big)^{p_{\boldsymbol{\gamma}}}. \tag{3.57}$$

The conclusion (3.45) follows by plugging (3.57) into (3.21). The theorem is proved.

□

3.3 Special Cases

In this section, Theorem 2 is applied to two special cases: balanced complete factorial q-way ANOVA models with main effects only and those with all the interaction effects. We also give an example in which the theorem can be applied to a fractional design.

3.3.1 Balanced Complete Factorial q-Way ANOVA Models with Main Effects

For a balanced q-way ANOVA model with main effects only, suppose the j-th factor has p_j levels ($j = 1, \ldots, q$) and each combination of levels has k replicates, then the model is the same as (3.3) with $m = q$ and $n = k \prod_{j=1}^{q} p_j$. The design matrices are

$$\boldsymbol{X}_0 = \mathbf{1}_n, \quad \boldsymbol{X}_j = \otimes_{i=1}^{m} \boldsymbol{U}_i^{(j)} \otimes \mathbf{1}_k,$$

for $j = 1, \ldots, q$, where $\boldsymbol{U}_i^{(j)}$ is $\boldsymbol{I}_{p_j}$ for $i = j$, and it is $\mathbf{1}_{p_i}$ when $i \neq j$. Clearly, the commutativity condition (3.28) is valid because

$$\boldsymbol{P}_i \boldsymbol{P}_j = \boldsymbol{P}_0 = \frac{1}{n} \boldsymbol{J}_n, \ \forall i \neq j = 0, \ldots, q. \tag{3.58}$$

The proposed prior is

$$\begin{aligned} \pi(\boldsymbol{\beta}_0, \sigma^2) &= \frac{1}{\sigma^2}, & (3.59) \\ \boldsymbol{\beta}_j \mid \sigma^2, g_j &\overset{indep}{\sim} N_{p_j}\big(\mathbf{0}, \sigma^2 \frac{p_j g_j}{n} \boldsymbol{I}_{p_j}\big). & (3.60) \end{aligned}$$

For the expression (3.46), for $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, we have:

$$(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}} = \begin{cases} \boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n, & if\ \boldsymbol{\gamma} = \{j\}, \\ \mathbf{0}_{n \times n}, & if\ |\boldsymbol{\gamma}| \geq 2. \end{cases} \tag{3.61}$$

Also, $u_{\{j\}} = -\frac{g_j}{1+g_j}$. Then

$$\boldsymbol{R} = (\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n) - \sum_{j=1}^{q} \frac{g_j}{1+g_j}(\boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n). \tag{3.62}$$

Similarly, we get

$$\boldsymbol{A}_{\boldsymbol{\gamma}} = \begin{cases} \boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n, & if\ \boldsymbol{\gamma} = \{j\}, \\ \boldsymbol{0}_{n\times n}, & if\ |\boldsymbol{\gamma}| \geq 2, \end{cases} \quad \text{and } p_{\boldsymbol{\gamma}} = \begin{cases} p_j - 1, & if\ \boldsymbol{\gamma} = \{j\}, \\ 0, & if\ |\boldsymbol{\gamma}| \geq 2. \end{cases} \tag{3.63}$$

Therefore, the marginal likelihood function under this situation is

$$m(\boldsymbol{y} \mid \boldsymbol{g}) = \frac{\Gamma\left(\frac{n-1}{2}\right) \prod_{j=1}^{q} (1+g_j)^{-\frac{p_j-1}{2}}}{\sqrt{n}\pi^{\frac{n-1}{2}} \left[\boldsymbol{y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{y} - \sum_{j=1}^{q} \frac{g_j}{1+g_j} \boldsymbol{y}'(\boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{y}\right]^{\frac{n-1}{2}}}. \tag{3.64}$$

Here, $\boldsymbol{y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{y}$ is the corrected total sum of squares, and for $j = 1, \ldots, q$, $\boldsymbol{y}'(\boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{y}$ is the sum of squares corresponding to the j-th factor.

Example 2. *(Two-way ANOVA model with main effects). Consider the model*

$$\boldsymbol{y} \;\; = \;\; \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{3.65}$$

where $\boldsymbol{y}$ *is* $n = p_1p_2k$ *dimensional column vector,* $\boldsymbol{X}_0 = \boldsymbol{1}_n$*,* $\boldsymbol{X}_1 = \boldsymbol{I}_{p_1} \otimes \boldsymbol{1}_{p_2} \otimes \boldsymbol{1}_k$*,* $\boldsymbol{X}_2 = \boldsymbol{1}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{1}_k$*,* $\boldsymbol{\beta}_1$ *and* $\boldsymbol{\beta}_2$ *are vectors of unknown main effects with dimensions* p_1 *and* p_2*, respectively, and* $\boldsymbol{\epsilon} \sim N_n\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n\big)$*. Since* $(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} = \frac{1}{p_2k}\boldsymbol{I}_{p_1}$ *and* $(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1} = \frac{1}{p_1k}\boldsymbol{I}_{p_2}$*, the priors we proposed are now*

$$\pi(\mu, \sigma^2) \;\; = \;\; \frac{1}{\sigma^2}, \tag{3.66}$$

$$(\boldsymbol{\beta}_1 \mid \sigma^2, g_1) \;\; \sim \;\; N_{p_1}\big(\boldsymbol{0}_{p_1\times 1}, \frac{g_1}{p_2k}\sigma^2\boldsymbol{I}_{p_1}\big), \tag{3.67}$$

$$(\boldsymbol{\beta}_2 \mid \sigma^2, g_2) \;\; \sim \;\; N_{p_2}\big(\boldsymbol{0}_{p_2\times 1}, \frac{g_2}{p_1k}\sigma^2\boldsymbol{I}_{p_2}\big). \tag{3.68}$$

The corresponding marginal likelihood is

$$\begin{aligned} & m(\boldsymbol{y} \mid g_1, g_2) \\ = \; & \frac{\Gamma(\frac{n-1}{2})}{\sqrt{n}\pi^{\frac{n-1}{2}} \prod_{i=1}^{2}(1+g_i)^{\frac{p_i-1}{2}}} \Big(SST - \frac{g_1}{1+g_1} SSA - \frac{g_2}{1+g_2} SSB \Big)^{-\frac{n-1}{2}}, \end{aligned} \tag{3.69}$$

where

$$SST = \boldsymbol{y}'\big(\boldsymbol{I}_n - \boldsymbol{P}_0\big)\boldsymbol{y} = \boldsymbol{y}'\Big(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\Big)\boldsymbol{y}, \tag{3.70}$$

$$SSA = \boldsymbol{y}'\big(\boldsymbol{P}_1 - \boldsymbol{P}_0\big)\boldsymbol{y} = \frac{1}{p_2 k}\boldsymbol{y}'\Big(\Big(\boldsymbol{I}_{p_1} - \frac{1}{p_1}\boldsymbol{J}_{p_1}\Big) \otimes \boldsymbol{J}_{p_2} \otimes \boldsymbol{J}_k\Big)\boldsymbol{y}, \tag{3.71}$$

$$SSB = \boldsymbol{y}'\big(\boldsymbol{P}_2 - \boldsymbol{P}_0\big)\boldsymbol{y} = \frac{1}{p_1 k}\boldsymbol{y}'\Big(\boldsymbol{J}_{p_1} \otimes \Big(\boldsymbol{I}_{p_2} - \frac{1}{p_2}\boldsymbol{J}_{p_2}\Big) \otimes \boldsymbol{J}_k\Big)\boldsymbol{y}. \tag{3.72}$$

They are the sum of squares corresponding to corrected total, factor A, and factor B, respectively.

The scales of the priors for $\boldsymbol{\beta}$'s now depend on the number of replicates, which might seem inappropriate. However, if we incorporate the hyper-prior (after adjusted for TESS) on g_1 and g_2, which is, for $j = 1, 2$,

$$g_j \overset{indep}{\sim} \textit{Inv-Gamma}\Big(1/2, n/(2p_j)\Big). \tag{3.73}$$

Let $g_j^* = p_j g_j / n$ for $j = 1, 2$, our proposed prior for $\boldsymbol{\beta}$'s will be equivalent to the following priors, for $j = 1, 2$,

$$(\boldsymbol{\beta}_j \mid \sigma^2, g_j^*) \overset{indep}{\sim} N_{p_j}\big(\mathbf{0}_{p_j \times 1}, g_j^* \sigma^2 \boldsymbol{I}_{p_j}\big), \tag{3.74}$$

$$g_j^* \overset{indep}{\sim} \textit{Inv-Gamma}\Big(1/2, 1/2\Big). \tag{3.75}$$

This coincides with the prior proposed independently by Rouder et al. (2012) that approached the problem from another perspective, and it is more reasonable since the scale of the prior does not depend on the number of replicates in the experiment now.

3.3.2 Balanced Completely Factorial q-Way ANOVA Models with All Interactions

For a q-way ANOVA model with all main and interaction effects, $m = 2^q - 1$ in (3.3). For $j = 1, \ldots, q$, we assume that the j-th factor has p_j levels. With k replicates at each combination of levels, there are $n = k \prod_{j=1}^{q} p_j$ obsersations in total. We use the subsets of $\{1, \ldots, q\}$ as the subscripts of the design matrices and covariate vectors. Then, the model (3.3) can be written as

$$\boldsymbol{y} = \boldsymbol{X}_{\emptyset}\boldsymbol{\beta}_{\emptyset} + \sum_{\emptyset \neq \tau \subset \{1,\ldots,q\}} \boldsymbol{X}_{\tau}\boldsymbol{\beta}_{\tau} + \boldsymbol{\epsilon}, \tag{3.76}$$

where $\boldsymbol{\beta}_\tau$ is a $p_\tau = \prod_{j \in \tau} p_j$ dimensional vector of regression coefficients, $\boldsymbol{X}_\tau$ is the $n \times p_\tau$ known design matrix, defined by

$$\boldsymbol{X}_\tau \;\; = \;\; \otimes_{i=1}^{q} \boldsymbol{V}_i^{(\tau)} \otimes \mathbf{1}_k.$$

Here, $\boldsymbol{V}_i^{(\tau)} = \boldsymbol{I}_{p_i}$ when $i \in \tau$ and $\boldsymbol{V}_i^{(\tau)} = \mathbf{1}_{p_i}$ when $i \notin \tau$. Clearly, $\boldsymbol{X}_\tau$ is of full column rank and the commutativity condition (3.28) is valid because

$$\boldsymbol{P}_\tau \boldsymbol{P}_{\tau^*} = \boldsymbol{P}_{\tau^*}\boldsymbol{P}_\tau = \boldsymbol{P}_{\tau \cap \tau^*}, \;\forall \tau, \tau^* \subseteq \{1, \ldots, q\}. \tag{3.77}$$

The prior is

$$\pi(\boldsymbol{\beta}_{\emptyset}, \sigma^2) \quad = \quad \frac{1}{\sigma^2}, \tag{3.78}$$

$$\boldsymbol{\beta}_\tau \mid \sigma^2, g_\tau \quad \stackrel{indep}{\sim} \quad N_{p_\tau}\big(\mathbf{0}, \sigma^2 \frac{g_\tau \prod_{j\in\tau} p_j}{n} \boldsymbol{I}_{p_\tau}\big). \tag{3.79}$$

In this case, the set of nonempty subscripts is $\boldsymbol{\Gamma}^* = \{\{1\}, \ldots, \{q\}, \{1,2\} \ldots, \{1, \ldots, q\}\}$ (equivalent to $\{1, \ldots, m\}$ in model (3.3)), and $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma} \neq \emptyset : \boldsymbol{\gamma} \subseteq \boldsymbol{\Gamma}^*\}$.

For each $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, define

$$\xi_{\boldsymbol{\gamma}} = \xi(\boldsymbol{\gamma}) = \bigcap_{\xi \in \boldsymbol{\gamma}} \xi. \tag{3.80}$$

We first show that for any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$,

$$\boldsymbol{A}_{\boldsymbol{\gamma}} \neq \mathbf{0} \Rightarrow \boldsymbol{\gamma} = \Big\{ \tau : \tau \supseteq \xi(\boldsymbol{\gamma}) \Big\}. \tag{3.81}$$

In fact, by the definition of $\xi(\boldsymbol{\gamma})$, $\forall \tau \in \boldsymbol{\gamma}$, $\xi(\boldsymbol{\gamma}) \subseteq \tau$. On the other hand, if $\exists \tau \supseteq \xi(\boldsymbol{\gamma})$ *s.t.* $\tau \notin \boldsymbol{\gamma}$, then $(\boldsymbol{I}_n - \boldsymbol{P}_\tau)\boldsymbol{P}_{\xi_{\boldsymbol{\gamma}}} = \mathbf{0}_{n\times n}$,so (3.81) holds. Therefore, $\forall \boldsymbol{A}_{\boldsymbol{\gamma}} \neq \mathbf{0}$,

$$\boldsymbol{A}_{\boldsymbol{\gamma}} = \boldsymbol{P}_{\xi_{\boldsymbol{\gamma}}} \prod_{\tau \subsetneq \xi_{\boldsymbol{\gamma}}} (\boldsymbol{I}_n - \boldsymbol{P}_\tau) = \prod_{\tau \subsetneq \xi_{\boldsymbol{\gamma}}} (\boldsymbol{P}_{\xi_{\boldsymbol{\gamma}}} - \boldsymbol{P}_{\tau \cap \xi_{\boldsymbol{\gamma}}}) = \prod_{j \in \xi_{\boldsymbol{\gamma}}} (\boldsymbol{P}_{\xi_{\boldsymbol{\gamma}}} - \boldsymbol{P}_{\xi_{\boldsymbol{\gamma}} \setminus \{j\}}), \tag{3.82}$$

so

$$p_{\boldsymbol{\gamma}} = \sum_{\tau \subseteq \xi_{\boldsymbol{\gamma}}} (-1)^{|\xi_{\boldsymbol{\gamma}}| - |\tau|} p_\tau. \tag{3.83}$$

In (3.45),

$$\prod_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \Big(1 + \sum_{j \in \boldsymbol{\gamma}} g_j \Big)^{-p_{\boldsymbol{\gamma}}/2} = \prod_{\emptyset \neq \xi \subseteq \{1,\ldots,q\}} \Big(1 + \sum_{\tau \supseteq \xi} g_\tau \Big)^{-\frac{1}{2} \sum_{\tau^* \subseteq \xi} (-1)^{|\xi| - |\tau^*|} p_{\tau^*}}. \tag{3.84}$$

Next, we need to calculate $\boldsymbol{R}$ in this case. In (3.46), since $\boldsymbol{P}_{\boldsymbol{\gamma}} = \prod_{\xi \in \boldsymbol{\gamma}} \boldsymbol{P}_\xi = \boldsymbol{P}_{\xi_{\boldsymbol{\gamma}}}$,

$$\sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{\phi}}) \boldsymbol{P}_{\boldsymbol{\gamma}} = \sum_{\emptyset \neq \tau \subseteq \{1,2,\ldots,q\}} \Bigg((\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{\phi}}) \boldsymbol{P}_\tau \sum_{\boldsymbol{\gamma} : \xi(\boldsymbol{\gamma}) = \tau} u_{\boldsymbol{\gamma}} \Bigg).$$

Lemma 3. *For nonempty* $\tau \subseteq \{1, 2, \ldots, q\}$, *define*

$$U_\tau = \sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : \xi(\boldsymbol{\gamma}) = \tau} u_{\boldsymbol{\gamma}}. \tag{3.85}$$

Then we have

$$\sum_{\tau^*\supseteq\tau} U_{\tau^*} = -1 + \frac{1}{1+\sum_{\tau^*\supseteq\tau} g_{\tau^*}}. \tag{3.86}$$

$$U_\tau = \begin{cases} \sum_{\tau^*\supseteq\tau} (-1)^{|\tau^*|-|\tau|} \frac{1}{1+\sum_{\tau^{**}\supseteq\tau^*} g_{\tau^{**}}}, & \text{if } \tau \subsetneq \{1,\dots,q\}, \\ -1 + \frac{1}{1+g_{\{1,\dots,q\}}}, & \text{if } \tau = \{1,\dots,q\}. \end{cases} \tag{3.87}$$

Proof. For (3.86), note that for any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$

$$\xi(\boldsymbol{\gamma}) \supseteq \tau \Leftrightarrow \boldsymbol{\gamma} \subseteq \{\tau^* : \tau^* \supseteq \tau\}. \tag{3.88}$$

Therefore, using Lemma 2(e), we get

$$\sum_{\tau^*\supseteq\tau} U_{\tau^*} = \sum_{\emptyset\neq\boldsymbol{\gamma}\subseteq\left\{\tau^*:\tau^*\supseteq\tau\right\}} u_{\boldsymbol{\gamma}} = -1 + \frac{1}{1+\sum_{\tau^*\supseteq\tau} g_{\tau^*}}. \tag{3.89}$$

For (3.87), we use mathematical induction. If $\tau = \{1,\dots,q\}$, (3.87) is exactly (3.86). If $\tau = \{1,\dots,q-1\}$, from (3.86), we have

$$U_{\{1,\dots,q\}} + U_{\{1,\dots,q-1\}} = -1 + \frac{1}{1+g_{\{1,\dots,q-1\}}+g_{\{1,\dots,q\}}}, \tag{3.90}$$

which implies that

$$U_{\{1,\dots,q-1\}} = \frac{1}{1+g_{\{1,\dots,q-1\}}+g_{\{1,\dots,q\}}} - \frac{1}{1+g_{\{1,\dots,q\}}}. \tag{3.91}$$

This proves that (3.87) holds for τ with $|\tau| = q-1$. Clearly, (3.86) implies a recursive formula,

$$U_\tau = -1 + \frac{1}{1+\sum_{\tau^*\supseteq\tau} g_{\tau^*}} - \sum_{\tau^*\supsetneq\tau} U_{\tau^*}. \tag{3.92}$$

Suppose (3.87) holds for τ with $|\tau| = k+1$, then

$$
\begin{aligned}
& U_{\{1,\ldots,k\}} \\
= \; & -1 + \frac{1}{1+\sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} - \sum_{\tau \supsetneq \{1,\ldots,k\}} U_\tau \\
= \; & \frac{1}{1+\sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} - \sum_{\tau \supsetneq \{1,\ldots,k\}} \left(\sum_{\tau^* \supseteq \tau} (-1)^{|\tau^*|-|\tau|} \frac{1}{1+\sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} \right) \\
= \; & \frac{1}{1+\sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} - \sum_{\tau^* \supsetneq \{1,\ldots,k\}} \left(\frac{1}{1+\sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} \left(\sum_{\tau : \tau^* \supseteq \tau \supsetneq \{1,\ldots,k\}} (-1)^{|\tau^*|-|\tau|} \right) \right) \\
= \; & \frac{1}{1+\sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} + \sum_{\tau^* \supsetneq \{1,\ldots,k\}} \left(\frac{1}{1+\sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} (-1)^{|\tau^*|-k} \right) \\
= \; & \sum_{\tau^* \supseteq \{1,\ldots,k\}} \left(\frac{1}{1+\sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} (-1)^{|\tau^*|-k} \right). \qquad (3.93)
\end{aligned}
$$

Therefore, (3.87) holds for τ with $|\tau| = k$. Repeat this procedure recursively, we can show that (3.87) holds for any nonempty $\tau \subsetneq \{1, \ldots, q\}$. The lemma is proved. $\square$

Theorem 3. *The marginal likelihood for an m-way ANOVA model (3.76) is*

$$
m(\boldsymbol{y} \mid \boldsymbol{g}) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{n}(\pi \boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n-1}{2}}} \prod_{\emptyset \neq \xi \subseteq \{1,\ldots,q\}} \left(1 + \sum_{\tau \supseteq \xi} g_\tau\right)^{-\frac{1}{2}\sum_{\tau^* \subseteq \xi} (-1)^{|\xi|-|\tau^*|} p_{\tau^*}}, \qquad (3.94)
$$

where

$$
\boldsymbol{R} = (\boldsymbol{I}_n - \boldsymbol{P}_\emptyset) + \sum_{\emptyset \neq \tau \subseteq \{1,2,\ldots,q\}} U_\tau (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{\phi}}) \boldsymbol{P}_\tau, \qquad (3.95)
$$

and U_τ is as defined in (3.87).

The next example is an illustration to the previous theorem.

Example 1 (Continued). *This model is a special case of (3.76) with $q = 2$. The* *corresponding marginal likelihood is*

$$m(\boldsymbol{y} \mid g_1, g_2, g_3) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{n}\pi^{\frac{n-1}{2}}}(1+g_1+g_3)^{-\frac{p_1-1}{2}}(1+g_2+g_3)^{-\frac{p_2-1}{2}}(1+g_3)^{-\frac{p_1p_2-p_1-p_2+1}{2}}$$
$$\left[SST - \frac{g_3}{1+g_3}SSAB - \frac{g_1+g_3}{1+g_1+g_3}SSA - \frac{g_2+g_3}{1+g_2+g_3}SSB\right]^{-\frac{n-1}{2}}, \quad (3.96)$$

where SST, SSA, and SSB are as defined in Example 2, and

$$\begin{aligned} SSAB &= \boldsymbol{y}'\big(\boldsymbol{P}_3 - \boldsymbol{P}_2 - \boldsymbol{P}_1 + \boldsymbol{P}_0\big)\boldsymbol{y} \\ &= \frac{1}{k}\boldsymbol{y}'\left(\left(\boldsymbol{I}_{p_1} - \frac{1}{p_1}\boldsymbol{J}_{p_1}\right)\otimes\left(\boldsymbol{I}_{p_2} - \frac{1}{p_2}\boldsymbol{J}_{p_2}\right)\otimes\boldsymbol{J}_k\right)\boldsymbol{y}. \end{aligned} \quad (3.97)$$

It is the sums of squares of the AB interaction.

3.3.3 Fractional Factorial Design

The two special cases introduced above are both complete factorial designs. The next example illustrates that Theorem 2 can be applied to a fractional design.

Example 3. *Consider the factorial design with 3 factors, where the factors have 2^J, 2^K, and 2^L levels, respectively (without loss of generality we assume that $J \geq K \geq L$). Consider the model,*

$$\boldsymbol{y} = \boldsymbol{1}_{2^{J+K}}\mu + \boldsymbol{X}_1\boldsymbol{\alpha} + \boldsymbol{X}_2\boldsymbol{\beta} + \boldsymbol{X}_3\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (3.98)$$

where

$$\boldsymbol{X}_1 = \boldsymbol{I}_{2^J}\otimes\boldsymbol{1}_{2^K},\ \boldsymbol{X}_2 = \boldsymbol{1}_{2^J}\otimes\boldsymbol{I}_{2^K},\ \boldsymbol{X}_3 = \boldsymbol{1}_{2^{J-L}}\otimes\begin{pmatrix}\boldsymbol{E}_1\\ \vdots \\ \boldsymbol{E}_{2^L}\end{pmatrix}. \quad (3.99)$$

In the expression above, $\boldsymbol{E}_1 = \boldsymbol{I}_{2^L}\otimes\boldsymbol{1}_{2^{K-L}}$, while for $j = 2, 3, \ldots, 2^L$, the first

$2^{K-L}(2^L+1-j)$ rows of $\boldsymbol{E}_j$ are the same as the last $2^{K-L}(2^L+1-j)$ rows of $\boldsymbol{E}_1$, and the last $2^{K-L}(j-1)$ rows of $\boldsymbol{E}_j$ are the same as the first $2^{K-L}(j-1)$ rows of $\boldsymbol{E}_1$. The projection matrices corresponding to $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$ are commutative, and the marginal likelihood can be calculated under (3.5) and (3.6).

3.4 On Commutativity of Projection Matrices

3.4.1 Orthogonal Arrays and Designs

The simple closed form expression of the marginal likelihood function (3.45) relies on (3.28) that the projection matrices are commutative under matrices multiplication,

$$\boldsymbol{P}_i\boldsymbol{P}_j = \boldsymbol{P}_j\boldsymbol{P}_i,\ \forall\ i, j.$$

Next, we discuss a class of models that satisfy this condition, for which we introduce the orthogonal arrays following Hedayat et al. (1999).

Definition 1. *An orthogonal array $OA(N, s_1 s_2 \dots s_K, t)$ is an $N \times K$ array, namely*

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \dots & & & \\ a_{N1} & a_{N2} & \dots & a_{NK} \end{pmatrix}, \tag{3.100}$$

where for $j = 1, 2, \dots, K$, $a_{ij} \in \{0, 1, \dots, s_j - 1\}$ ($\forall i = 1, \dots, N$), such that in any $N \times t$ subarray every possible t-tuple (sequence of t elements) occurs an equal number of times as a row.

An orthogonal array suggests a design. For example, consider the orthogonal array in (3.100). We can construct a factorial design with N design points and K factors, where the j-th factor (for $j = 1, \ldots, K$) has s_j levels denoted with $\{0, 1, \ldots, s_j - 1\}$. For the i-th design point ($i = 1, \ldots, N$), we assign level a_{ij} for the j-th factor ($j = 1, \ldots, K$).

The design matrices for the main effects and the interaction effects can be constructed. For the orthogonal array in (3.100), denote the design matrix for the main effect of the j-th factor with $\boldsymbol{X}_j$ (for $j = 1, \ldots, K$). $\boldsymbol{X}_j$ is an $N \times s_j$ matrix , and it can be constructed using the j-th column of the orthogonal array, $(a_{1j}, \ldots, a_{Nj})^T$. The $(i, a_{ij} + 1)$-th entry of $\boldsymbol{X}_j$ is equal to 1 for $i = 1, 2, \ldots, N$, and the other entries of $\boldsymbol{X}_j$ are equal to 0. For the interaction effects, we can combine two or more columns into one if we treat the combination of levels from two or more factors as one level, For example, if we consider the j_1-th and j_2-th columns ($j_1 \neq j_2$), $(a_{ij_1}, a_{ij_2}) \in A \equiv \{(x, y) : x \in \{0, 1, \ldots, s_{j_1} - 1\}, y \in \{0, 1, \ldots, s_{j_2} - 1\}\}$ for $i = 1, 2, \ldots, N$. There are $s_{j_1} \times s_{j_2}$ elements in A, so there exists a bijection $f : A \mapsto B$, where $B \equiv \{0, 1, \ldots, s_{j_1}s_{j_2} - 1\}$, then we can get a new column $(f(a_{1j_1}, a_{1j_2}), \ldots, f(a_{Nj_1}, a_{Nj_2}))'$ whose elements are from B. Using this new column, we can construct the $N \times s_{j_1}s_{j_2}$ design matrix $\boldsymbol{X}_{j_1j_2}$ for the interaction effect of the j_1-th and j_2-th factors.

Example 4. *For an* $OA(16, 2^5, 4)$ *(transposed):*

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \tag{3.101}$$

consider the design corresponding to this orthogonal array, let the 5 factors be A, B, C, D, and E, each with 2 levels. Then the matrix corresponding to the main effect of

A is

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}^T, \tag{3.102}$$

and the matrix corresponding to the interaction effect of A and B is

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}^T. \tag{3.103}$$

3.4.2 Orthogonal Arrays and Commutativity Condition

Next, we show that when $t = 2$ for an orthogonal array, the commutativity condition (3.28) is satisfied for the model with main effects only.

Theorem 4. *For an orthogonal array $OA(N, s_1 s_2 \ldots s_K, t)$ with $t = 2$, let the design matrices for the main effects $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$ be defined as above, and let*

$$\boldsymbol{P}_j = \boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\boldsymbol{X}_j',\ j = 1, \ldots, K.$$

For $j_1, j_2 \in \{1, 2, \ldots, K\}, j_1 \neq j_2$,

$$(\boldsymbol{P}_{j_1} - \frac{1}{N}\boldsymbol{J}_N)(\boldsymbol{P}_{j_2} - \frac{1}{N}\boldsymbol{J}_N) \;\; = \;\; \boldsymbol{0}, \tag{3.104}$$

and therefore,

$$\boldsymbol{P}_{j_1}\boldsymbol{P}_{j_2} - \boldsymbol{P}_{j_2}\boldsymbol{P}_{j_1} \;\; = \;\; \boldsymbol{0}. \tag{3.105}$$

Proof. We first notice that for each row of $\boldsymbol{X}_j$, only one entry is equal to 1, and the other entries are 0's. Whereas for each column of $\boldsymbol{X}_j$, N/s_j entries equal to 1, and the others are 0's. Therefore,

$$\boldsymbol{X}'_{j_l}\boldsymbol{X}_{j_l} = \frac{N}{s_{j_l}}\boldsymbol{I}_{s_{j_l}}, l = 1, 2, \tag{3.106}$$

and thus

$$\boldsymbol{P}_{j_1}\boldsymbol{P}_{j_2} = \frac{s_{j_1}s_{j_2}}{N^2}\boldsymbol{X}_{j_1}\boldsymbol{X}'_{j_1}\boldsymbol{X}_{j_2}\boldsymbol{X}'_{j_2}. \tag{3.107}$$

Further, notice that the (k,l)-th element in $\boldsymbol{X}'_{j_1}\boldsymbol{X}_{j_2}$ (for $k = 1, \ldots, s_{j_1}, l = 1, \ldots, s_{j_2}$) is equal to the number of design points that have the k-th level for the j_1-th factor and the l-th level for the j_2-th factor, so it is equal to $N/s_{j_1}s_{j_2}$. Therefore,

$$\begin{aligned}
\boldsymbol{X}'_{j_1}\boldsymbol{X}_{j_2} &= \frac{N}{s_{j_1}s_{j_2}}\boldsymbol{1}_{s_{j_1}\times s_{j_2}} \\
&= \frac{N}{s_{j_1}s_{j_2}}\boldsymbol{1}_{s_{j_1}\times 1}\boldsymbol{1}_{1\times s_{j_2}}, &&(3.108)\\
\boldsymbol{P}_{j_1}\boldsymbol{P}_{j_2} &= \frac{1}{N}(\boldsymbol{X}_{j_1}\boldsymbol{1}_{s_{j_1}\times 1})(\boldsymbol{1}_{1\times s_{j_2}}\boldsymbol{X}'_{j_2}) \\
&= \frac{1}{N}\boldsymbol{1}_{N\times 1}\boldsymbol{1}_{1\times N} \\
&= \frac{1}{N}\boldsymbol{J}_N. &&(3.109)
\end{aligned}$$

While

$$\begin{aligned}
\boldsymbol{P}_{j_1}\frac{1}{N}\boldsymbol{J}_N &= \frac{s_{j_1}}{N^2}\boldsymbol{X}_{j_1}\boldsymbol{X}'_{j_1}\boldsymbol{1}_{N\times 1}\boldsymbol{1}_{1\times N} \\
&= \frac{s_{j_1}}{N^2}\boldsymbol{X}_{j_1}\frac{N}{s_{j_1}}\boldsymbol{1}_{s_{j_1}\times 1}\boldsymbol{1}_{1\times N} \\
&= \frac{1}{N}\boldsymbol{J}_N, &&(3.110)
\end{aligned}$$

and since $\frac{1}{N}\boldsymbol{J}_N$ is symmetric,

$$\frac{1}{N}\boldsymbol{J}_N\boldsymbol{P}_{j_1} \quad = \quad \frac{1}{N}\boldsymbol{J}_N. \tag{3.111}$$

Similar result holds for $\boldsymbol{P}_{j_2}$, and therefore,

$$\begin{aligned}
& (\boldsymbol{P}_{j_1} - \frac{1}{N}\boldsymbol{J}_N)(\boldsymbol{P}_{j_2} - \frac{1}{N}\boldsymbol{J}_N) \\
= \ & \boldsymbol{P}_{j_1}\boldsymbol{P}_{j_2} - \frac{1}{N}\boldsymbol{J}_N\boldsymbol{P}_{j_2} - \boldsymbol{P}_{j_1}\frac{1}{N}\boldsymbol{J}_N - \frac{1}{N}\boldsymbol{J}_N\frac{1}{N}\boldsymbol{J}_N \\
= \ & \boldsymbol{0}.
\end{aligned}$$

The theorem is proved. □

In general, when $t \geq 2$ for an orthogonal array, we can include in the model some of the interaction effects, and the commutativity condition (3.28) still holds. To show this result, we need to first modify the notations. For $A \subseteq \{1, \ldots, K\}$, let $\boldsymbol{X}_A$ be the design matrix for the interaction of the factors in A, and let $\boldsymbol{P}_A$ be the corresponding projection matrix.

Theorem 5. *Consider an orthogonal array $OA(N, s_1 s_2 \ldots s_K, t)$, where $t \geq 2$. For two sets $A, B \subseteq \{1, \ldots, K\}$, where $|A \cup B|$, the number of elements in $A \cup B$, is less than or equal to t,*

$$\boldsymbol{P}_A\boldsymbol{P}_B = \boldsymbol{P}_B\boldsymbol{P}_A. \tag{3.112}$$

Proof. If $A \cap B = \phi$, then we can combine the columns in A into one column and the columns in B into another column. The array consisting of these two columns is an orthogonal array with $t = 2$, and the result follows from Theorem 4.

If $A \cap B \neq \phi$, we can prove that

$$(\boldsymbol{P}_A - \boldsymbol{P}_{A\cap B})(\boldsymbol{P}_B - \boldsymbol{P}_{A\cap B}) = \boldsymbol{0}. \tag{3.113}$$

First, we notice that $\boldsymbol{X}_{A\cap B}$ is in the column space of $\boldsymbol{X}_A$. Therefore,

$$\begin{aligned}\boldsymbol{P}_A\boldsymbol{P}_{A\cap B} &= \boldsymbol{P}_{A\cap B}, &(3.114)\\ \boldsymbol{P}_{A\cap B}\boldsymbol{P}_A &= \boldsymbol{P}_{A\cap B}. &(3.115)\end{aligned}$$

Similar results hold between $\boldsymbol{P}_{A\cap B}$ and $\boldsymbol{P}_B$, so we only need to show

$$\boldsymbol{P}_A\boldsymbol{P}_B = \boldsymbol{P}_{A\cap B}. \tag{3.116}$$

Similar to the proof of Theorem 4,

$$\begin{aligned}\boldsymbol{X}_A'\boldsymbol{X}_A &= \frac{N}{\prod_{j\in A} s_j}\boldsymbol{I}_{\prod_{j\in A} s_j}, &(3.117)\\ \boldsymbol{X}_B'\boldsymbol{X}_B &= \frac{N}{\prod_{j\in B} s_j}\boldsymbol{I}_{\prod_{j\in B} s_j}. &(3.118)\end{aligned}$$

Therefore,

$$\boldsymbol{P}_A\boldsymbol{P}_B = \frac{\prod_{j\in A} s_j \prod_{j\in B} s_j}{N^2}\boldsymbol{X}_A\boldsymbol{X}_A'\boldsymbol{X}_B\boldsymbol{X}_B'. \tag{3.119}$$

For $i_1, i_2, l \in \{1,\ldots,N\}$, the (i_1,l)-th element of $\boldsymbol{X}_A\boldsymbol{X}_A'$ is 1 if the i_1-th and the l-th design points have the same levels for the factors in A, and it is zero otherwise. Similarly, the (l,i_2)-th element of $\boldsymbol{X}_B\boldsymbol{X}_B'$ is 1 if the i_2-th and the l-th design points have the same levels for the factors in B, and it is zero otherwise. Therefore, the (i_1,i_2)-th element in $\boldsymbol{X}_A\boldsymbol{X}_A'\boldsymbol{X}_B\boldsymbol{X}_B'$ is the total number of design points that have the same levels in A as the i_1-th design point and the same levels in B as the i_2-th design point. It is equal to 0 if the i_1-th and i_2-th design points have different levels in $A\cap B$, and when the i_1-th and i_2-th design points have the same levels in $A\cap B$, it is equal to $N/\prod_{j\in A\cup B} s_j$. Therefore,

$$\boldsymbol{X}_A\boldsymbol{X}_A'\boldsymbol{X}_B\boldsymbol{X}_B' = \frac{N}{\prod_{j\in A\cup B} s_j}\boldsymbol{X}_{A\cap B}\boldsymbol{X}_{A\cap B}'. \tag{3.120}$$

Thus,

$$\boldsymbol{P}_A\boldsymbol{P}_B \;=\; \frac{\prod_{j\in A\cap B} s_j}{N}\boldsymbol{X}_{A\cap B}\boldsymbol{X}_{A\cap B}' = \boldsymbol{P}_{A\cap B}. \tag{3.121}$$

The theorem is proved. □

Corollary 1. *Consider the design for an orthogonal array $OA(N, s_1s_2\ldots s_K, t)$, where $t \geq 2$. For $T \subset \Gamma = \Big\{\{1\}, \{2\}, \ldots, \{K\}, \{1,2\}, \ldots, \{1,2,\ldots,K\}\Big\}$, such that $\forall A, B \in T$, $|A\cup B| \leq t$. If we include every interaction in T in the model (i.e., include X_A, $\forall A \in T$), condition (3.28) still holds.*

Specifically, T could be the collection of all the subsets of $\{1,\ldots,K\}$ that have no more than $\lfloor \frac{t}{2}\rfloor$ elements, and under this situation, the model include all the interactions up to order $\lfloor \frac{t}{2}\rfloor$.

From Theorem 4 and Corollary 1 we can see that, if we are not required to include all the interaction effects in a model, the orthogonal arrays suggest fractional factorial designs where the commutativity condition (3.28) holds. However, it should be noticed that such designs are balanced up to a certain level. Condition (3.28) no longer holds for unbalanced designs. Consider the following example.

Example 5. *Let*

$$\boldsymbol{X}_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}^T, \boldsymbol{X}_2 = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}^T,$$

then

$$\begin{aligned}
\boldsymbol{P}_1\boldsymbol{P}_2 &= \begin{pmatrix} \frac{1}{9}\mathbf{1}_3 & \frac{1}{6}\mathbf{1}_3 & \frac{1}{6}\mathbf{1}_3 & \frac{1}{9}\mathbf{1}_3 & \frac{1}{9}\mathbf{1}_3 & \frac{1}{6}\mathbf{1}_3 & \frac{1}{6}\mathbf{1}_3 \\ \frac{1}{6}\mathbf{1}_4 & \frac{1}{8}\mathbf{1}_4 & \frac{1}{8}\mathbf{1}_4 & \frac{1}{6}\mathbf{1}_4 & \frac{1}{6}\mathbf{1}_4 & \frac{1}{8}\mathbf{1}_4 & \frac{1}{8}\mathbf{1}_4 \end{pmatrix}, \\
\boldsymbol{P}_2\boldsymbol{P}_1 &= (\boldsymbol{P}_1\boldsymbol{P}_2)^T \neq \boldsymbol{P}_1\boldsymbol{P}_2.
\end{aligned}$$

3.5 Computation

In Theorems 1 and 2, the closed form expression of the marginal likelihood is derived given $\boldsymbol{g}$. However, if a hyper-prior on $\boldsymbol{g}$ is considered, finding the marginal likelihood usually involves a multidimensional integration that is analytically intractable. Numerical integration can be applied when $\boldsymbol{g}$ is either 1 or 2 dimensional, but it is not applicable when $\boldsymbol{g}$ has 3 or higher dimensions as the posterior will be highly concentrated. Different approximation techniques such as the Laplace approximation can also be implemented, but they cannot provide accurate values of the Bayes factors when the sample size is small. In this section, the inverse-gamma prior on $\boldsymbol{g}$ is considered as discussed in Section 3.2, and we give a procedure to obtain the Bayes factors by applying the Savage-Dickey density ratio (Dickey 1971, Verdinelli & Wasserman 1995).

First, the Bayes factor between two nested models is considered. The full and the reduced models are assumed to be, respectively,

$$M_F \;:\; \boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_R\boldsymbol{\beta}_R + \cdots + \boldsymbol{X}_{R+F}\boldsymbol{\beta}_{R+F} + \boldsymbol{\epsilon}, \quad (3.122)$$

$$M_R \;:\; \boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_R\boldsymbol{\beta}_R + \boldsymbol{\epsilon}, \quad (3.123)$$

where the settings on $\boldsymbol{y}$, $\boldsymbol{X}_j$, $\boldsymbol{\beta}_j$, and $\boldsymbol{\epsilon}$ are the same as those of (3.3). Note that $\boldsymbol{X}_0\boldsymbol{\beta}_0$ is the common part of all models under consideration. Denote the proposed prior for M_F and M_R as π_F and π_R, respectively. The Bayes factor between M_F and

M_R is

$$B_{FR} = m_F/m_R, \tag{3.124}$$

where m_F and m_R are the marginal likelihood functions corresponding to the two models.

Consider a transformation for M_F and π_F:

$$\boldsymbol{\beta}_j^* = \begin{cases} \boldsymbol{\beta}_j, & \text{for } j = 0, \ldots, R, \\ \frac{1}{\sqrt{g_j}\sigma}\boldsymbol{\beta}_j, & \text{for } j = R+1, \ldots, R+F. \end{cases} \tag{3.125}$$

Then the transformed model is

$$M_F^* \quad : \quad \boldsymbol{y} \sim N_n\Big(\sum_{j=0}^{R} \boldsymbol{X}_j\boldsymbol{\beta}_j^* + \sum_{j=R+1}^{R+F} \sqrt{g_j\sigma^2}\boldsymbol{X}_j\boldsymbol{\beta}_j^*, \sigma^2\boldsymbol{I}_n\Big) \tag{3.126}$$

with the prior,

$$\pi_F^*(\boldsymbol{\beta}_j^* \mid \sigma^2, g_j) \quad = \quad \begin{cases} \pi_F(\boldsymbol{\beta}_j^* \mid \sigma^2, g_j), & \text{for} \quad j = 1, \ldots, R, \\ N_{p_j}\big(\boldsymbol{\beta}_j^*; \boldsymbol{0}, (\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}\big), & \text{for} \quad j = R+1, \ldots, R+F, \end{cases} \tag{3.127}$$

$$\pi_F^*(g_j) \quad = \quad \pi_F(g_j), \text{ for } j = 1, \ldots, R+F, \tag{3.128}$$

$$\pi_F^*(\beta_0^*, \sigma^2) \quad = \quad \pi_F(\beta_0^*, \sigma^2). \tag{3.129}$$

Let m_F^* denote the marginal likelihood under M_F^* and π_F^*. Clearly, $m_F = m_F^*$ and the priors of $\boldsymbol{\beta}_j^*$ does not depend on (σ^2, g_j), for $j = R+1, \ldots, R+F$. Also, let M_R^* have the same likelihood function as M_R, and let the corresponding prior π_R^* be the same as π_R except for auxiliary variables $g_{R+1}, \ldots, g_{R+F}$ together with the priors $\pi_R^*(g_j) = \pi_F(g_j)$ for $j = R+1, \ldots, R+F$. The corresponding marginal likelihood $m_R^* = m_R$ since the prior of $(g_{R+1}, \ldots, g_{R+F})$ is proper.

The Bayes factor $B_{FR} = m_F/m_R$ is equal to m_F^*/m_R^*, which is the Bayes factor for testing M_F^* against M_R^* under the priors π_F^* and π_R^*, respectively. Under π_F^*, $\boldsymbol{\beta}_j^*$ $(j = R+1, \ldots, R+F)$ is independent of the other parameters, so the Savage-Dickey density ratio leads to

$$B_{FR} = \frac{\pi_F^*(\boldsymbol{\beta}_{R+j}^* = \mathbf{0}, j = 1, \ldots, F)}{\pi_F^*(\boldsymbol{\beta}_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y})} = \frac{\prod_{j=R+1}^{R+F} \frac{|\boldsymbol{X}_j'\boldsymbol{X}_j|^{1/2}}{(2\pi)^{p_j/2}}}{\pi_F^*(\boldsymbol{\beta}_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y})}. \quad (3.130)$$

The denominator $\pi_F^*(\beta_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y})$ does not have a closed form expression, but if we have a posterior sample $(\boldsymbol{\beta}_0^*, \ldots, \boldsymbol{\beta}_{R+F}^*, g_1, \ldots, g_{R+F}, \sigma^2)^{(i)}$ $(i = 1, \ldots, N)$ from $\pi_F^*(\cdot \mid \boldsymbol{y})$, then

$$\begin{aligned} & \pi_F^*(\beta_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y}) \\ \approx \; & \frac{1}{N} \sum_{i=1}^{N} \pi_F^*\big(\beta_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y}, (\boldsymbol{\beta}_0^*, \ldots, \boldsymbol{\beta}_R^*, g_1, \ldots, g_{R+F}, \sigma^2)^{(i)}\big). \quad (3.131) \end{aligned}$$

The remaining problem is sampling from $\pi_F^*(\cdot \mid \boldsymbol{y})$ and finding $\pi_F^*(\beta_{R+j}^* = \mathbf{0}, j = 1, \ldots, F \mid \boldsymbol{y}, \cdot)$.

For posterior sampling, note that the sample from $\pi_F(\cdot \mid \boldsymbol{y})$ can be transformed into the sample from $\pi_F^*(\cdot \mid \boldsymbol{y})$ by (3.125). Furthermore, the posterior sample from $\pi_F(\cdot \mid \boldsymbol{y})$ can be easily obtained via Gibbs sampling (Gelfand & Smith 1990). In the same way as (3.14), (3.16), and (3.17), we define $\boldsymbol{\beta}_F$, $\boldsymbol{X}_F$, and $\boldsymbol{M}_F$ for M_F. We also let $\tilde{\boldsymbol{\beta}}_F = \big(\boldsymbol{X}_F'\boldsymbol{X}_F + \boldsymbol{M}_F\big)^{-1}\boldsymbol{X}_F'\boldsymbol{y}$, then the full conditional posteriors for $\pi_F(\cdot \mid \boldsymbol{y})$ are given as follows

$$\begin{aligned} \boldsymbol{\beta}_F \mid \cdots &\sim N\big(\tilde{\boldsymbol{\beta}}_F, \sigma^2(\boldsymbol{X}_F'\boldsymbol{X}_F + \boldsymbol{M}_F)^{-1}\big), \\ \sigma^2 \mid \cdots &\sim \text{Inv-Gamma}\left(\frac{n + \sum_{j=1}^{R+F} p_j}{2}, \frac{(\boldsymbol{y} - \boldsymbol{X}_F\boldsymbol{\beta}_F)'(\boldsymbol{y} - \boldsymbol{X}_F\boldsymbol{\beta}_F) + \boldsymbol{\beta}_F'\boldsymbol{M}_F\boldsymbol{\beta}_F}{2}\right), \\ g_j \mid \cdots &\sim \text{Inv-Gamma}\Big(\frac{1+p_j}{2}, nb_j + \frac{\boldsymbol{\beta}_j'\boldsymbol{X}_j'\boldsymbol{X}_j\boldsymbol{\beta}_j}{2\sigma^2}\Big), \text{ for } j = 1, \ldots, R+F. \end{aligned}$$

The full conditional posterior of $\boldsymbol{\beta}_T^* = (\boldsymbol{\beta}_{R+1}^{*\prime}, \ldots, \boldsymbol{\beta}_{R+F}^{*\prime})'$ under M_F^* and π_F^* is

$$(\boldsymbol{\beta}_T^* \mid \boldsymbol{y}, \cdot) \sim N\Big(\frac{1}{\sigma}\big(\boldsymbol{L}_T'\boldsymbol{L}_T + \boldsymbol{M}_T\big)^{-1}\boldsymbol{L}_T'\Big(\boldsymbol{y} - \sum_{j=0}^{R}\boldsymbol{X}_j\boldsymbol{\beta}_j^*\Big), \big(\boldsymbol{L}_T'\boldsymbol{L}_T + \boldsymbol{M}_T\big)^{-1}\Big), \quad (3.132)$$

where $\boldsymbol{L}_T = \big(\sqrt{g_{R+1}}\boldsymbol{X}_{R+1}, \ldots, \sqrt{g_{R+F}}\boldsymbol{X}_{R+F}\big)$ and $\boldsymbol{M}_T = diag\big(\boldsymbol{X}_{R+1}'\boldsymbol{X}_{R+1}, \ldots, \boldsymbol{X}_{R+F}'\boldsymbol{X}_{R+F}\big)$. After further simplification, it can be shown that

$$\begin{aligned} B_{RF} = \frac{1}{B_{FR}} \;\approx\; & \frac{1}{N}\sum_{i=1}^{N}\Big|\boldsymbol{I}_n + \sum_{j=R+1}^{R+F} g_j^{(i)}\boldsymbol{P}_j\Big|^{\frac{1}{2}} \exp\Big\{-\frac{1}{2(\sigma^2)^{(i)}}\Big(\boldsymbol{y} - \sum_{j=0}^{R}\boldsymbol{X}_j\boldsymbol{\beta}_j^{*(i)}\Big)' \\ & \Big[\boldsymbol{I}_n - \big(\boldsymbol{I}_n + \sum_{j=R+1}^{R+F} g_j^{(i)}\boldsymbol{P}_j\big)^{-1}\Big]\Big(\boldsymbol{y} - \sum_{j=0}^{R}\boldsymbol{X}_j\boldsymbol{\beta}_j^{*(i)}\Big)\Big\}. \end{aligned} \quad (3.133)$$

Finally, for two non-nested models M_{R_1} and M_{R_2}, a global full model M_F that contains all the variables in M_{R_1} and M_{R_2} can be considered. Then $B_{R_1R_2} = B_{R_1F}/B_{R_2F}$ where B_{R_1F} and B_{R_2F} can be computed using the method introduced in this section.

3.6 Simulation

In the previous sections, the closed form expression of the marginal likelihood given $\boldsymbol{g}$ is obtained, and how the Bayes factors can be computed with hyper inverse-gamma priors on g's is discussed. The choice of hyperparameters in the priors of g's is still a problem. In this section, simulation studies are conducted to examine 2 choices of the hyperparameters. Following the idea of Zellner-Siow prior, b's are first set to 1/2 in the prior of $\boldsymbol{g}$. Then each component of $\boldsymbol{g}$ has the same hyper-prior Inv-Gamma$(1/2, n/2)$, and this method is denoted as 'IZS' in Tables 3.5–3.19. Alternatively, the choice of b's can be adjusted according to 'The Effective Sample Size' (Berger et al. 2010) of the corresponding parameters, which is denoted as 'TESS'. Sun et al. (2012) showed that for one-way ANOVA models, priors adjusted according to 'TESS' will lead to Bayes factors that are consistent in model selection. For comparison, the generalized

Jeffreys-Zellner-Siow prior proposed by Bayarri & García-Donato (2007) is used. This prior is also designed to deal with the linear models with non-full-column-rank design matrices, and it is referred to as 'BG'. For the three methods, the Bayes factors are interpreted in half-units on the $\log_{10}$ scale (see Table 3.1) as suggested by Jeffreys (1961), Kass & Raftery (1995).

3.6.1 Simulation Case 1

In this simulation study, the Bayes factors are calculated under different priors to test for the interaction effects in 2-way ANOVA models. Specifically, the model in Example 1 (denoted as M_4) and its sub-model without the interaction effect (denoted as M_3) are considered. Data is generated from these two models, and then we examine whether the three approaches can correctly attribute the data to its origin. For 'TESS', the hyper-priors on g_1, g_2, and g_3 are Inv-Gamma$(1/2, p_2k/2)$, Inv-Gamma$(1/2, p_1k/2)$, and Inv-Gamma$(1/2, k/2)$, respectively.

Data Generation: p_1 and p_2 are set as 5, 10, 20, or 50, and k is set to 5 or 20. Without loss of generality, μ is set as 0, and σ^2 is set as 1 (see Lemma 4 for an explanation). $\boldsymbol{\beta}_1$ is generated from $N(\mathbf{0}, \tilde{g}_1\sigma^2\boldsymbol{I}_{p_1})$, $\boldsymbol{\beta}_2$ is generated from $N(\mathbf{0}, \tilde{g}_2\sigma^2\boldsymbol{I}_{p_2})$, $\boldsymbol{\beta}_3$ is generated from $N(\mathbf{0}, \tilde{g}_3\sigma^2\boldsymbol{I}_{p_1p_2})$, and $\boldsymbol{\epsilon}$ is generated from $N(\mathbf{0}, \sigma^2\boldsymbol{I}_{p_1p_2k})$, where $\tilde{g}_1$ and $\tilde{g}_2$ are fixed at 1, and $\tilde{g}_3$ assumes values 0, 0.1, and 0.3. The simulated data $\boldsymbol{y}$ is then calculated according to M_4. Note that when $\tilde{g}_3 = 0$, $\boldsymbol{y}$ is from M_3, the model without interaction, whereas when $\tilde{g}_3 \neq 0$, $\boldsymbol{y}$ is from M_4, the model with interaction, and greater values of $\tilde{g}_3$ correspond to greater interaction effects in the model. This procedure is repeated 100 times under each combination of p_1, p_2, k, and $\tilde{g}_3$. For each set of the simulated $\boldsymbol{y}$, $\ln(B_{43})$ is calculated under the three priors, and the means and the standard deviations of $\ln(B_{43})$ under each prior are summarized in Tables 3.5–3.10.

Lemma 4. *The Bayes factors calculated with the* 3 *priors do not change if* μ *is changed or if* $\boldsymbol{\beta}_i$ *(*$i = 1, 2, 3$*) and* $\boldsymbol{\epsilon}$ *are changed proportionally when generating the data* $\boldsymbol{y}$*.*

Proof. B_{43} calculated from the three methods depends on the data $\boldsymbol{y}$ only through SSA/SST, SSB/SST, and $SSAB/SST$, which are invariant to the change of μ and the proportional change of $\boldsymbol{\beta}_i$ ($i = 1, 2, 3$) and $\boldsymbol{\epsilon}$. The lemma is proved. $\square$

Interpretation: When $\tilde{g}_3 = 0$ (*i.e.* under model M_3), methods 'BG' and 'IZS' give negative $\ln(B_{43})$ with large absolute values, which supports M_3. 'TESS' also gives the desirable results when p_1, p_2, and k are large enough, but when the sample size is small, $\ln(B_{43})$ is close to 0 on average, which means that the evidence supporting M_3 is weak. When p_1, p_2, and k are all small, note that $\ln(B_{43})$ calculated from 'TESS' could be positive occasionally, which supports the wrong model. This undesirable situation improves for greater values of p_1 and p_2.

When $\tilde{g}_3 = 0.1$, there exists a weak interaction effect, then 'TESS' clearly outperforms its competitors. Methods 'BG' and 'IZS' fail to detect the interactions even when p_1, p_2, and k are large. 'TESS' only fails when p_1, p_2, and k are all small, whereas when the sample size increases, it discovers the interaction. Furthermore, 'TESS' always yields the desirable results when $\tilde{g}_3 = 0.3$, whereas 'IZS' still fails for small sample size, and 'BG' could fail even when the sample size is large.

To summarize, 'TESS' outperforms the other two methods in comparing models M_3 and M_4 when the sample size is moderate to large, for it can better separate cases with interactions and those without interactions. However, when the sample size is small, 'TESS' is biased towards M_4 in the sense that sometimes it cannot identify the data from M_3, whereas 'IZS' and 'BG' are biased towards M_3 as they cannot detect the interaction effect when it is weak. In terms of model selection consistency, 'TESS' leads to consistent Bayes factor, whereas the Bayes factors calculated from the other two priors are not always consistent.

3.6.2 Simulation Case 2

In this part, a similar simulation study is performed to examine the main effect in a 2-way ANOVA model. The models being considered are M_3 and

$$M_1 : \boldsymbol{y} = \mathbf{1}_{p_1p_2k}\mu + \boldsymbol{I}_{p_1} \otimes \mathbf{1}_{p_2} \otimes \mathbf{1}_k\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}. \tag{3.134}$$

For 'TESS', the same priors as in Section 3.6.1 are applied to different components of $\boldsymbol{g}$.

Data Generation: Without loss of generality, μ is set as 0, and σ^2 is set as 1. p_1 and p_2 are set to 3, 5, 10, 20, or 50, and k is set to 3, 5, or 20. $\boldsymbol{\beta}_1$ is generated from $N_{p_1}(\mathbf{0}, \tilde{g}_1\sigma^2\boldsymbol{I}_{p_1})$, $\boldsymbol{\beta}_2$ is generated from $N_{p_2}(\mathbf{0}, \tilde{g}_2\sigma^2\boldsymbol{I}_{p_2})$, and $\boldsymbol{\epsilon}$ is generated from $N_n(\mathbf{0}, \sigma^2\boldsymbol{I}_{p_1p_2k})$, where $\tilde{g}_1$ is fixed at 1, and $\tilde{g}_2$ assumes values 0, 0.1, and 0.3. Under each setting, we calculate the data $\boldsymbol{y}$ according to M_3 and repeat for 100 times . $\ln(B_{31})$ is calculated for the 100 groups of data under the three priors, and the means and the standard deviations of $\ln(B_{31})$ are summarized under each prior in Tables 3.11–3.19.

Interpretation: When M_1 is the true model (*i.e.* $\tilde{g}_2 = 0$), all three methods suggest the correct model with negative averages of $\ln(B_{31})$. The absolute values of $\ln(B_{31})$ from the three methods are close for small sample sizes, and they increase as the sample size increases.

When $\tilde{g}_2 \neq 0$, 'TESS' gives better results than the other two methods as it yields greater values of $\ln(B_{31})$ and it always leads to the correct choice of model except for the smallest sample size. Also note that for small $\tilde{g}_2$, if p_1 and k are fixed at small values, $\ln(B_{31})$ given by methods 'BG' and 'IZS' sometimes decreases as p_2 increases, which suggests that the Bayes factor is inconsistent. Since 'BG' and 'IZS' do not always lead to consistent Bayes factor, our recommended prior for comparing models M_3 and M_1 is again 'TESS'. However, it should be used with cautious since it could select the wrong model when p_1, p_2, and k are small and the effect of $\boldsymbol{\beta}_2$ is weak.

3.7 Analyzing Real Data

In this section, the proposed methods 'IZS' and 'TESS' are applied to the linear models on two real data sets to explore the small sample properties. For both data sets, all possible linear models with the given covariates are considered. The intercept term is always included in the models as the common parameter, and the other covariates are centered at zero in order to reduce confounding effects. Since no grouping information is available for the covariates, each covariate is considered a group by itself. We use the uniform prior on all possible models when finding the posterior probabilities. The multiplicity adjusting prior presented in Scott & Berger (2010) is a possible alternative but is not applied here since the focus of this study is to discuss the effect of the Bayes factor.

3.7.1 Hald Data

The first example is the Hald data, which is available in the R library 'monomvn'. This data set is taken from Wood et al. (1932) aiming to analyze the effect of composition of cement on the heat evolved during setting and hardening. There are 13 observations with a response variable Y and 4 covariates X_j ($j = 1, \ldots, 4$). Table 3.2 includes a description of these variables. The Hald data has been commonly used in literature on model selection. See, for example, George & McCulloch (1993), Hald (1952), Deltell (2011). For both methods being considered, the posterior inclusion probabilities of the covariates are also listed in Table 3.2.

In Table 3.3, the 6 most probable models and their posterior probabilities are given according to 'IZS'. The results obtained by 'IZS' and 'TESS' are similar to those from George & McCulloch (1993), Deltell (2011). However, in terms of both the most probable models and the inclusion probabilities, our methods favor the models with more covariates, which agrees with the observation from the simulation studies.

3.7.2 Crime Data

The second example is the US crime data of Vandaele (1978), which is also commonly analyzed in literature about Bayesian model selection (Raftery et al. 1997, Fernández et al. 2001, Liang et al. 2008, Deltell 2011). It is available as dataset 'UScrime' in the 'MASS' library of R. The response variable is the crime rate for 47 US states in 1960, and 15 crime-related and demographic variables are included as explanatory variables, which leads to $2^{15} = 36768$ possible linear models. A detailed description of the 15 variables is listed in Table 3.4. Following other literature, logarithm are taken on all the variables except the indicator variable 'So'. Table 3.4 also summarizes the posterior inclusion probabilities of the 15 covariates using 'IZS' and 'TESS'. The proposed methods are again giving results similar to the previous literature for most covariates. For example, covariates 'Ed' and 'Ineq', which have the highest inclusion probabilities in literature, are also chosen with the highest probabilities from our method. However, two highly correlated covariates 'Po1' and 'Po2' have inflated inclusion probabilities, which might suggest a potential issue of the proposed methods in dealing with correlated covariates or a necessity of considering such covariates as one group. Also note that the inclusion probabilities for some covariates change dramatically from 'IZS' to 'TESS', further justification is needed on which one is more appropriate for linear models.

3.8 Comments

In this chapter, we propose a modification of Zellner's g-prior for the Bayes factors of linear models. This prior is designed to overcome the difficulty of Zellner's g-prior for models with non-full-column-rank design matrices such as ANOVA models, and it can also bring more flexibility to the priors. We calculate the marginal likelihood functions for the proposed prior, and a simpler form of the marginal likelihood is derived under the commutativity condition of the projection matrices. The commutativity condition is shown to hold for a class of fractional factorial designs using the tool of orthogonal arrays. As illustrations to the general result, the marginal likelihood functions of the balanced q-way ANOVA models with either main effects only or with all interaction effects are calculated using this closed form expression. Examples are given for balanced 2-way ANOVA models and a 3-factor fractional factorial design. Next, the approach for computing the Bayes factors with hyper-priors on $\boldsymbol{g}$ is given. The simulation studies show that, to acquire consistent Bayes factor, the hyper-prior on $\boldsymbol{g}$ should be chosen according to 'The Effective Sample Size' of the corresponding parameters. Out of the three methods being compared, this proposed prior performs the best in model comparison for 2-way ANOVA models. Finally, the proposed methods 'IZS' and 'TESS' are applied to 2 real data sets and are shown to yield satisfactory results.

Table 3.1: Jeffreys' Scale of Evidence for Bayes Factor

$\ln(B_{ij})$	$\log_{10}(B_{ij})$	Interpretation
$\ln(B_{ij}) < -4.61$	$\log_{10}(B_{ij}) < -2$	Decisive evidence for M_j
$-4.61 < \ln(B_{ij}) < -2.30$	$-2 < \log_{10}(B_{ij}) < -1$	Strong evidence for M_j
$-2.30 < \ln(B_{ij}) < -1.15$	$-1 < \log_{10}(B_{ij}) < -1/2$	Substantial evidence for M_j
$-1.15 < \ln(B_{ij}) < 0$	$-1/2 < \log_{10}(B_{ij}) < 0$	Poor evidence for M_j
$0 < \ln(B_{ij}) < 1.15$	$0 < \log_{10}(B_{ij}) < 1/2$	Poor evidence for M_i
$1.15 < \ln(B_{ij}) < 2.30$	$1/2 < \log_{10}(B_{ij}) < 1$	Substantial evidence for M_i
$2.30 < \ln(B_{ij}) < 4.61$	$1 < \log_{10}(B_{ij}) < 2$	Strong evidence for M_i
$4.61 < \ln(B_{ij})$	$2 < \log_{10}(B_{ij})$	Decisive evidence for M_i

Table 3.2: Description and Posterior Inclusion Probabilities of Variables: Hald Data.

Variable	Description	IZS	TESS
Y	Heat evolved (calories/gram)		
X_1	Percentage weight in clinkers of tricalcium aluminate	0.997	0.995
X_2	Percentage weight in clinkers of tricalcium silicate	0.821	0.836
X_3	Percentage weight in clinkers of tetracalcium aluminoferrile	0.405	0.458
X_4	Percentage weight in clinkers of β-di-calcium silicate	0.689	0.693

Table 3.3: 6 Most Probable Models (IZS) and Posterior Probabilities: Hald Data.

Model	IZS	TESS
X_1, X_2, X_4	0.316	0.299
X_1, X_2	0.215	0.186
X_1, X_2, X_3, X_4	0.192	0.227
X_1, X_3, X_4	0.115	0.105
X_1, X_2, X_3	0.096	0.121
X_1, X_4	0.063	0.057

Table 3.4: Description and Posterior Inclusion Probabilities of Variables: US Crime Data.

Variable	Description	IZS	TESS
y(response)	Rate of crime in a particular category per head of population		
M	Percentage of males aged 14-24	0.825	0.835
So	Indicator variable for a Southern state	0.309	0.344
Ed	Mean years of schooling	0.980	0.921
Po1	Police expenditure in 1960	0.764	0.791
Po2	Police expenditure in 1959	0.681	0.659
LF	Labor force participation rate	0.147	0.287
M.F	Number of males per 1000 females	0.132	0.338
Pop	State population	0.284	0.395
NW	Number of non-whites per 1000 people	0.636	0.775
U1	Unemployment rate of urban males 14-24	0.189	0.301
U2	Unemployment rate of urban males 35-39	0.584	0.557
GDP	Gross domestic product per head	0.365	0.491
Ineq	Income inequality	0.998	0.992
Prob	Probability of imprisonment	0.847	0.849
Time	Average time in state prisons	0.314	0.364

Table 3.5: Mean and Standard Deviation of $\ln(B_{43})$: $k = 5$, $\tilde{g}_3$=0

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$	BG	-16.55(3.62)	-39.26(6.68)	-96.21(7.92)	-323.55(17.15)
	IZS	-10.93(3.21)	-20.71(6.1)	-37.77(7.43)	-95.67(13.89)
	TESS	-2.16(1.71)	-2.84(2.18)	-4.01(1.92)	-6.4(2.5)
$p_2 = 10$	BG		-104.35(6.92)	-263.22(15.39)	-870.31(33.54)
	IZS		-43.74(6.84)	-86.59(10.61)	-217.77(20.74)
	TESS		-4.18(1.81)	-5.62(2.46)	-11.42(6.31)
$p_2 = 20$	BG			-675.51(25.61)	-2153.1(74.22)
	IZS			-178.79(15.24)	-458.48(24.08)
	TESS			-9.62(4.98)	-33.99(11.55)
$p_2 = 50$	BG				-6642.31(187.23)
	IZS				-1197.58(37.89)
	TESS				-106.45(16.66)

Table 3.6: Mean and Standard Deviation of $\ln(B_{43})$: $k = 20$, $\tilde{g}_3$=0

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$	BG	-27.35(3.21)	-63.96(4.61)	-150.03(7.02)	-463.97(18.94)
	IZS	-17.43(4.49)	-38.06(7.04)	-87.91(10.46)	-229.44(17.19)
	TESS	-5.45(2.18)	-7.67(2.39)	-10.25(2.58)	-13.85(3.68)
$p_2 = 10$	BG		-161.04(8.58)	-386.56(14.74)	-1190.85(38.91)
	IZS		-90.27(10.16)	-190(14.89)	-491.38(25.31)
	TESS		-10.64(2.58)	-13.79(3.07)	-20.81(5.26)
$p_2 = 20$	BG			-936.36(29.47)	-2844.62(63.39)
	IZS			-394.86(21.67)	-1011.27(28.52)
	TESS			-18.61(4.87)	-45.13(10.7)
$p_2 = 50$	BG				-8420.25(179.55)
	IZS				-2587.34(49.03)
	TESS				-134.88(16.3)

Table 3.7: Mean and Standard Deviation of $\ln(B_{43})$: $k = 5$, $\tilde{g}_3$=0.1

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
	BG	-12.88(4.52)	-32.25(7.58)	-77.7(10.82)	-273.98(20.44)
$p_2 = 5$	IZS	-7.44(4.09)	-12.82(6.39)	-23.96(9.94)	-69.85(17.69)
	TESS	-0.48(2.12)	0.45(2.77)	2.21(3.37)	7.12(5.64)
	BG		-84.85(10.36)	-228.66(17.78)	-761.14(42.78)
$p_2 = 10$	IZS		-25.32(9.29)	-54.21(14.91)	-142.13(23.17)
	TESS		2.43(3.15)	5.8(4.69)	16.94(7.04)
	BG			-591.57(34.66)	-1936.44(70.54)
$p_2 = 20$	IZS			-114.58(18.77)	-297.68(27.71)
	TESS			13.14(6.52)	36.03(10.26)
	BG				-6121.23(169.81)
$p_2 = 50$	IZS				-745.49(46.43)
	TESS				96.02(16.92)

Table 3.8: Mean and Standard Deviation of $\ln(B_{43})$: $k = 20$, $\tilde{g}_3$=0.1

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
	BG	-11.85(8.13)	-28.91(12.94)	-80.88(17.33)	-272.56(35.22)
$p_2 = 5$	IZS	-0.63(8.48)	1.82(14.77)	-8.9(21.02)	-27.64(34.63)
	TESS	5.97(6.09)	15.76(9.89)	30.26(11.98)	89.58(22.76)
	BG		-86.47(17.98)	-218.3(28.82)	-766.51(58.7)
$p_2 = 10$	IZS		-11.04(21.18)	-16.08(31.21)	-58.32(49.78)
	TESS		31.57(11.85)	76.06(18.74)	194.71(32.1)
	BG			-596.66(44.3)	-1951.88(85.57)
$p_2 = 20$	IZS			-46.74(40.61)	-125.54(60.99)
	TESS			155.04(26.54)	405.85(45.33)
	BG				-6126.18(232.68)
$p_2 = 50$	IZS				-321.85(103.81)
	TESS				1069.93(73.58)

Table 3.9: Mean and Standard Deviation of $\ln(B_{43})$: $k = 5$, $\tilde{g}_3$=0.3

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$	BG	-7.27(6.81)	-17.5(9.25)	-48.51(14.76)	-190.49(23.28)
	IZS	-0.84(7.66)	2.64(10.1)	10.46(14.91)	26.15(24.09)
	TESS	3.3(4.25)	7.77(5.46)	17.52(9.25)	47.02(13.24)
$p_2 = 10$	BG		-52.47(14.81)	-155.34(22.05)	-570.65(45.92)
	IZS		9.66(15.27)	19.15(22.43)	47.79(38.9)
	TESS		18.32(8.6)	38.42(12.23)	111.96(22.54)
$p_2 = 20$	BG			-444.27(37.14)	-1545.31(88.51)
	IZS			35.79(28.74)	80.01(53.67)
	TESS			85.28(18.26)	238.99(37.43)
$p_2 = 50$	BG				-5082.08(189.05)
	IZS				197.14(78.77)
	TESS				643.92(57.7)

Table 3.10: Mean and Standard Deviation of $\ln(B_{43})$: $k = 20$, $\tilde{g}_3$=0.3

	Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$	BG	19.13(15.1)	27.53(24.51)	62.62(32.97)	61.3(62.91)
	IZS	36(17.71)	75.67(30.96)	185.54(41.84)	429.36(70.75)
	TESS	33.19(14.2)	65.58(23.96)	158.63(34.71)	408.24(68.42)
$p_2 = 10$	BG		59.11(35.2)	62.99(50.26)	-9.46(97.12)
	IZS		176.61(42.62)	347.87(62.37)	877.33(105.27)
	TESS		161.12(36.98)	335.35(52.05)	910.12(95.98)
$p_2 = 20$	BG			15.11(75.02)	-341.17(150.88)
	IZS			684.94(92.77)	1764.42(158.66)
	TESS			718.01(81.76)	1917.42(138.84)
$p_2 = 50$	BG				-2078.71(248.04)
	IZS				4352.43(234.15)
	TESS				4901.15(229.07)

Table 3.11: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-1.76(1.34)	-1.97(1.08)	-2.06(1.11)	-2.09(0.92)	-2.86(2.59)
	IZS	-2.32(1.3)	-2.9(1.1)	-3.55(1.12)	-4.34(1.1)	-8.66(3.77)
	TESS	-1.51(1.14)	-2.01(0.96)	-2.62(1.01)	-3.24(0.84)	-4.04(0.94)
$p_2 = 5$	BG	-4.03(1.51)	-4.29(1.34)	-4.6(1.67)	-4.77(1.34)	-6.23(2.88)
	IZS	-4.65(1.46)	-5.6(1.29)	-6.85(1.69)	-8.35(1.51)	-13.81(4.21)
	TESS	-2.4(1.17)	-3.1(1.07)	-4.12(1.47)	-5.27(1.28)	-6.97(1.56)
$p_2 = 10$	BG	-10.25(2.99)	-11.36(2.6)	-12.33(2.41)	-13.33(2.28)	-16.63(3.64)
	IZS	-10.31(2.31)	-12.55(2.34)	-15.51(2.33)	-18.61(2.3)	-26.51(4.87)
	TESS	-3.68(1.63)	-4.98(1.78)	-6.95(1.98)	-9.22(1.9)	-13.05(2.05)
$p_2 = 20$	BG	-23.08(4.13)	-26.17(3.97)	-30.41(3.27)	-33.25(3.54)	-40.63(4.78)
	IZS	-22.93(3.58)	-27.57(3.38)	-33.89(3.16)	-40.53(3.55)	-52.1(5.46)
	TESS	-5.46(1.93)	-7.42(2.21)	-10.74(2.13)	-14.97(2.51)	-22.52(3.39)
$p_2 = 50$	BG	-67.84(6.59)	-80.11(6.4)	-93.2(6.42)	-106.12(6.17)	-126.15(8.32)
	IZS	-71.22(7.11)	-84.39(6.79)	-100.17(6.6)	-116.83(6.72)	-143.96(7.98)
	TESS	-8.85(2.6)	-12.87(2.78)	-19.81(4.15)	-30.76(4.48)	-47.64(5.28)

Table 3.12: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-2.12(1.27)	-2.31(1.07)	-2.44(1.14)	-2.61(0.93)	-3.37(2.71)
	IZS	-2.7(1.2)	-3.33(1.08)	-4(1.16)	-4.97(1.14)	-9.37(4.02)
	TESS	-1.81(1.09)	-2.39(0.99)	-3.02(1.08)	-3.76(0.83)	-4.49(1.02)
$p_2 = 5$	BG	-4.99(1.47)	-5.14(1.53)	-5.55(1.54)	-5.68(1.4)	-7.2(3.12)
	IZS	-5.61(1.38)	-6.59(1.44)	-7.87(1.48)	-9.43(1.54)	-14.85(4.38)
	TESS	-3.08(1.19)	-3.86(1.26)	-4.97(1.37)	-6.16(1.34)	-7.94(1.48)
$p_2 = 10$	BG	-12.28(2.48)	-13.25(2.56)	-14.6(2.23)	-15.44(2.57)	-18.9(3.91)
	IZS	-12.53(2.25)	-14.65(2.14)	-17.74(2.03)	-20.74(2.66)	-28.33(4.87)
	TESS	-4.93(1.68)	-6.31(1.79)	-8.58(1.8)	-10.96(2.19)	-15.31(2.11)
$p_2 = 20$	BG	-27.28(3.54)	-31.6(3.75)	-34.96(4.15)	-38.37(3.5)	-45.75(4.5)
	IZS	-27.75(3.46)	-32.42(3.2)	-38.27(3.81)	-45.33(3.66)	-57.39(6.11)
	TESS	-7.53(2.03)	-9.81(2.4)	-13.52(2.93)	-19.1(2.85)	-26.69(3.34)
$p_2 = 50$	BG	-80.8(7.94)	-92.13(6.83)	-105.24(6.96)	-118.65(6.87)	-137.51(8.39)
	IZS	-83.36(7.52)	-95.95(7.21)	-112.5(7.2)	-130.27(6.87)	-155.93(8.43)
	TESS	-12.23(3.14)	-17.52(3.81)	-27.92(5.01)	-39.02(6.09)	-57.74(6.91)

Table 3.13: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
	BG	-3.55(0.98)	-3.52(1.22)	-3.86(1.03)	-3.8(0.89)	-4.57(2.75)
$p_2 = 3$	IZS	-4.2(0.93)	-4.62(1.21)	-5.51(1)	-6.38(1.06)	-9.98(3.9)
	TESS	-3.16(0.91)	-3.56(1.19)	-4.43(0.97)	-5.05(0.86)	-5.88(1.11)
	BG	-7.45(1.64)	-8.05(1.56)	-8.12(1.48)	-8.41(1.44)	-9.94(2.9)
$p_2 = 5$	IZS	-8.28(1.58)	-9.34(1.44)	-10.46(1.44)	-12.19(1.48)	-16.91(4.48)
	TESS	-5.3(1.51)	-6.27(1.4)	-7.35(1.38)	-8.88(1.36)	-10.9(1.28)
	BG	-18.07(2.67)	-19.67(2.27)	-20.71(2.3)	-21.57(2.48)	-24.52(3.79)
$p_2 = 10$	IZS	-18.58(2.29)	-20.64(2.08)	-23.83(2.16)	-26.94(2.33)	-34.79(4.37)
	TESS	-9.27(2.08)	-11.02(1.9)	-13.96(2.05)	-16.71(2.32)	-20.94(2.38)
	BG	-40.03(3.8)	-44.09(3.89)	-47.87(3.54)	-51.41(3.97)	-57.81(5.46)
$p_2 = 20$	IZS	-40.51(3.86)	-45.06(3.35)	-51.7(3.37)	-58.78(3.75)	-70.95(4.99)
	TESS	-15.14(3.03)	-18.95(2.91)	-24.25(3.32)	-30.59(3.35)	-39.39(3.29)
	BG	-113.86(7.58)	-126.34(7.35)	-139.44(5.96)	-151.78(6.09)	-172.38(7.59)
$p_2 = 50$	IZS	-116.62(6.89)	-131.09(7.19)	-147.81(6.89)	-164(7)	-190.6(7.55)
	TESS	-30.53(4.76)	-39.55(5.96)	-54.58(6.11)	-70.31(6.8)	-91.7(6.47)

Table 3.14: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
	BG	-0.87(1.63)	-0.36(2.53)	0.87(3.31)	3.36(6.08)	14.54(14.77)
$p_2 = 3$	IZS	-1.44(1.68)	-1.26(2.58)	-0.59(3.43)	1.26(6.37)	8.86(14.66)
	TESS	-0.76(1.47)	-0.56(2.31)	0.09(3.11)	1.96(5.87)	11.61(13.41)
	BG	-2.66(2.18)	-1.71(3.65)	1.23(5.03)	5.51(8.28)	21.89(19.43)
$p_2 = 5$	IZS	-3.3(2.16)	-2.92(3.71)	-0.99(5.07)	2.16(8.48)	14.13(19.49)
	TESS	-1.31(1.76)	-0.8(3.2)	1.16(4.59)	4.44(7.79)	18.75(18.45)
	BG	-6.97(3.81)	-5.35(4.82)	-0.16(7.06)	10.33(12.32)	48.27(27.31)
$p_2 = 10$	IZS	-7.31(3.73)	-6.7(4.58)	-3.04(7.15)	5.02(12.4)	37.46(27.72)
	TESS	-1.49(2.84)	-0.36(3.7)	3.73(6.24)	12.2(11.4)	47.31(25.92)
	BG	-15.11(6.61)	-13.86(7.53)	-4.82(11.15)	20.27(18.77)	97.39(40.55)
$p_2 = 20$	IZS	-14.9(5.85)	-14.46(6.84)	-8.61(10.5)	12.82(18.93)	84.3(40.6)
	TESS	-0.31(3.87)	1.85(5)	9.43(8.89)	31.74(16.64)	106.02(38.27)
	BG	-49.45(9.85)	-42.63(10.87)	-24.19(16.26)	37.55(30.37)	227.73(72.56)
$p_2 = 50$	IZS	-51.55(10.69)	-47.51(12.12)	-32.3(17.47)	25.04(30.22)	209.53(72.12)
	TESS	1.3(5.1)	9.71(7.39)	31.08(13)	91.06(26.01)	280.07(67.34)

Table 3.15: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-0.89(2.13)	0.15(3.03)	1.53(4.72)	5.38(8.14)	23.41(22.5)
	IZS	-1.46(2.13)	-0.79(3.07)	0(4.84)	3.2(8.37)	17.27(22.88)
	TESS	-0.69(1.96)	-0.02(2.85)	0.78(4.58)	4.01(7.92)	20.39(21.67)
$p_2 = 5$	BG	-2.39(2.97)	-0.35(4.46)	2.88(6.28)	15.88(15.11)	38.29(37.34)
	IZS	-3.06(2.91)	-1.63(4.53)	0.59(6.33)	12.47(15.33)	29.79(37.58)
	TESS	-0.87(2.56)	0.61(4.13)	2.92(5.94)	14.63(14.6)	35.22(36.04)
$p_2 = 10$	BG	-5.35(5.11)	-4.25(6.43)	6.45(11.69)	24.32(20.04)	88.43(55.78)
	IZS	-5.91(4.96)	-5.39(6.27)	3.28(11.7)	18.99(20.09)	77.8(55.95)
	TESS	0.34(4.03)	1.51(5.37)	10.36(10.72)	26.41(19)	87.46(54.01)
$p_2 = 20$	BG	-14.69(7.51)	-8.61(9.36)	6.6(14.66)	52.44(31.49)	172.78(67.7)
	IZS	-13.91(7.19)	-9.88(9.25)	3.01(14.89)	44.94(31.89)	159.35(68.09)
	TESS	2.33(5.2)	7.73(7.2)	22.04(12.96)	64.32(29.62)	182.06(65.56)
$p_2 = 50$	BG	-45.22(11.82)	-35.66(15.57)	9.32(28.39)	112.31(46.98)	437.77(112.72)
	IZS	-49.12(12.57)	-39.57(15.78)	0.3(29.13)	100.14(47.22)	420.25(112.29)
	TESS	8.74(7.16)	22.66(11.31)	66.93(24.02)	167.9(43.21)	491.93(107.32)

Table 3.16: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	2.31(6.63)	4.7(8.32)	12.39(15.44)	38.29(35.92)	91.07(96.48)
	IZS	1.71(6.7)	3.74(8.36)	10.8(15.54)	35.94(36.23)	84.77(96.68)
	TESS	2.58(6.52)	4.62(8.2)	11.64(15.31)	36.79(35.75)	88.1(95.85)
$p_2 = 5$	BG	3.07(8.28)	9.03(13.28)	32.22(26.82)	68.73(49.69)	172.2(123.37)
	IZS	2.54(8.25)	7.7(13.2)	29.95(26.96)	65.15(49.84)	163.76(124.26)
	TESS	5.02(7.91)	10.23(12.82)	32.31(26.49)	67.55(49.23)	169.14(122.42)
$p_2 = 10$	BG	4.83(11.81)	23.12(20.79)	58.92(43.33)	143.06(80.3)	388.47(184.21)
	IZS	4.53(11.81)	21.81(20.91)	55.73(43.41)	137.59(80.24)	378.58(184.74)
	TESS	12(10.94)	29.26(19.95)	63.35(42.35)	145.23(79.11)	388.34(183.25)
$p_2 = 20$	BG	11.6(18.07)	43.46(32.62)	127.12(55.66)	317.8(112.97)	826.28(278.66)
	IZS	11.05(18.25)	42.04(32.42)	123.3(55.49)	310.69(113.18)	815.39(279)
	TESS	30.3(16.32)	61.8(30.31)	143.25(53.53)	330.26(110.88)	836.96(276.06)
$p_2 = 50$	BG	27.88(29.23)	105.7(49.07)	341.01(95.71)	770.93(207.88)	2161.1(452.02)
	IZS	23.2(28.31)	100.89(49.07)	332.62(96.23)	760.61(207.77)	2144.63(451.8)
	TESS	89.87(24.52)	168.79(44.77)	401.13(91.94)	829.38(203.18)	2214.98(447.69)

Table 3.17: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	0.1(2.68)	1.38(3.6)	5.54(6.92)	13.61(13.75)	38.76(36.7)
	IZS	-0.42(2.74)	0.54(3.73)	4.3(7.23)	12(14.41)	33.97(38.49)
	TESS	0.15(2.44)	1.05(3.36)	4.55(6.63)	11.85(13.31)	34.64(34.83)
$p_2 = 5$	BG	0.67(4.44)	4.35(6.16)	8.66(9.19)	25.91(21.64)	68.36(44.63)
	IZS	0(4.45)	3.17(6.19)	6.59(9.37)	23.07(22.2)	62.04(45.66)
	TESS	1.48(3.83)	4.53(5.52)	8.04(8.55)	23.96(20.84)	63.55(42.83)
$p_2 = 10$	BG	0.2(6.57)	6.15(9.44)	22.4(14.47)	54.02(29.4)	156.77(68.73)
	IZS	-0.38(6.35)	4.7(9.3)	19.56(14.54)	49.16(29.82)	147.49(69.41)
	TESS	3.92(5.14)	9.18(8.03)	23.92(13.24)	53.37(28.05)	152.13(66.6)
$p_2 = 20$	BG	-4.34(8.87)	10.74(12.61)	42.79(23.6)	116.87(50.21)	325.76(102.46)
	IZS	-3.97(8.41)	9.96(12.62)	39.09(23.71)	109.97(50.51)	313.8(101.75)
	TESS	7.31(6.26)	21.11(10.53)	50.95(21.27)	121.72(47.46)	325.95(98.33)
$p_2 = 50$	BG	-14.98(16.06)	17.54(20.96)	102.48(39.88)	268.2(68.44)	854.12(179.44)
	IZS	-17.48(15.55)	13.18(20.63)	94.86(39.71)	258.07(69.64)	836.89(178.1)
	TESS	22.45(10.72)	54.67(16.34)	138.18(35.23)	301.73(64.33)	881.77(173.24)

Table 3.18: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	2.05(4.59)	4.24(6.49)	6.88(8.41)	22.64(22.8)	57.02(52.44)
	IZS	1.49(4.65)	3.37(6.61)	5.49(8.63)	20.96(23.49)	52.2(53.54)
	TESS	2.03(4.33)	3.86(6.2)	5.96(8.16)	20.92(22.39)	53.22(50.7)
$p_2 = 5$	BG	2.74(6.22)	6.96(9.1)	19.5(18.71)	51.93(35.57)	120.5(85.54)
	IZS	2.05(6.25)	5.69(9.22)	17.43(18.94)	49.17(36.25)	113.71(86.86)
	TESS	3.67(5.63)	7.31(8.55)	18.83(18.01)	49.89(34.9)	115.7(83.63)
$p_2 = 10$	BG	3.67(7.75)	13.72(11.77)	44.08(26.91)	94.62(47.24)	292.8(128.82)
	IZS	3.07(7.65)	12.41(11.85)	41.19(26.98)	89.61(47.53)	283.68(128.95)
	TESS	7.86(6.58)	17.26(10.67)	45.71(25.52)	93.97(46.05)	287.72(125.63)
$p_2 = 20$	BG	9.3(12.97)	32.11(20.39)	82.44(34.48)	208.26(71.9)	569.8(177.36)
	IZS	9.27(12.84)	30.91(20.25)	78.6(34.45)	201.47(72.14)	559.1(178.25)
	TESS	20.71(10.59)	42.64(18.08)	91.12(32.4)	213.43(69.56)	570.12(174.13)
$p_2 = 50$	BG	17.69(21.4)	68.57(29.74)	209.26(56.95)	522.87(112.7)	1456.2(286)
	IZS	14.49(21.59)	63.46(29.96)	201.86(57.84)	511.76(112.67)	1438.75(286.26)
	TESS	55.89(16.97)	107.4(25.43)	246.51(53.29)	556.17(108.64)	1485.18(281.32)

Table 3.19: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
	BG	12.29(14.54)	20.74(21.33)	48.39(51.63)	103.76(97.45)	248.89(204.9)
$p_2 = 3$	IZS	11.73(14.57)	19.87(21.41)	47.05(52.01)	102.04(98.1)	243.67(205.8)
	TESS	12.37(14.27)	20.44(21.04)	47.38(51.34)	102.07(97.05)	245.11(203.1)
	BG	23.92(17.61)	47.27(31.85)	87.35(51.58)	192.6(137.1)	457.19(313.5)
$p_2 = 5$	IZS	23.22(17.54)	45.99(31.82)	85.17(51.78)	189.57(137.9)	451.74(314.5)
	TESS	24.97(17.02)	47.61(31.2)	86.7(51.08)	190.77(136.6)	452.8(311.6)
	BG	54.14(33.18)	93.95(44.43)	219.38(98.92)	444.46(208.5)	1233.9(489.4)
$p_2 = 10$	IZS	53.61(33.27)	92.47(44.47)	216.47(98.98)	439.6(208.7)	1224.9(489.6)
	TESS	58.6(32.01)	97.53(43.32)	221.15(97.52)	443.85(207.0)	1229.1(487.0)
	BG	108.69(47.82)	205.48(83.05)	445.2(139.2)	907.64(261.3)	2391.0(714.9)
$p_2 = 20$	IZS	108.22(48.3)	204.69(83.2)	441.8(138.7)	900.64(261.9)	2378.8(715.4)
	TESS	121.05(45.71)	217.47(80.47)	454.07(136.9)	912.93(259.2)	2391.5(711.8)
	BG	263.2(73)	514.34(106.0)	1135.5(212.8)	2403.5(439.5)	6198.0(1029.7)
$p_2 = 50$	IZS	261.92(73.38)	510.2(105.5)	1127.6(213.3)	2392.5(439.9)	6181.4(1030.0)
	TESS	306.24(68.37)	555.64(102.1)	1173.8(208.7)	2437.6(435.5)	6226.6(1025.0)

Chapter 4

Consistency of Bayes Factors for 2-Way ANOVA Models with Main Effects

4.1 Introduction

In Chapter 3, a modified Zellner's g-prior is proposed for the ANOVA models. Two choices of the hyper-prior on g's are considered. First, Inv-Gamma$(1/2, n/2)$ prior can be chosen for each g_j as in the original Zellner-Siow prior, which we refer to as 'Independent-Zellner-Siow' (IZS) prior. Alternatively, n can be replaced by 'The Effective Sample Size' in the inverse gamma priors, which we call 'TESS' method. The simulation studies justified the use of 'TESS' as it gives consistent Bayes factors whereas 'IZS' does not under certain circumstances. However, this justification is only empirical rather than theoretical. In this chapter, a more formal discussion of the Bayes factor consistency is given.

Consistency is a criterion that is often used to evaluate model comparison methods. It means that assuming one of the models under comparison is true, then this true model will be selected if enough data is observed. However, traditional methods such as p-values, C_p, and AIC all fail to guarantee consistency, whereas Bayesian model selection is often consistent (Berger & Pericchi 2001). In terms of Bayes factor, consistency means that for the true model M_T and any other model M_i under comparison, when sample size increases, the Bayes factor of M_i versus M_T

$$B_{iT} \to 0. \tag{4.1}$$

For the consistency of Bayes factors with Zellner's g-prior, Fernández et al. (2001) investigated different choices of fixed g, Liang et al. (2008) studied both the fixed g-priors and the mixtures of g-priors with three families of proper priors for g when the model dimension is fixed, Maruyama & George (2011) used the beta-prime prior on g and applied it to the case when the number of parameters can be greater than the number of observations, and Guo & Speckman (2009) studied both the reference prior and the usual proper priors for g when the model dimension is allowed to increase with the number of observations. Sun et al. (2012) showed that a modified Zellner-Siow prior can be used for both fixed and random effects in one-way ANOVA models, and they discussed the conditions for consistency.

There is also a large amount of literature concerning Bayes factor consistency with other priors. Berger et al. (2003) discussed the consistency issue of the Bayes factor in one-way ANOVA models with the variance assumed to be known, and they showed that the multivariate Cauchy prior and the smooth Cauchy prior have different consistency behavior when the number of levels is large. García-Donato & Sun (2007) showed the consistency of the Bayes factor for one-way random effects model for fixed and growing model dimension. Casella et al. (2009) proved that in the class of normal linear models with a finite number of regressors, the Bayes factor procedure is consistent for a wide class of priors including intrinsic priors. Moreno et al. (2010) studied the consistency in models where the number of parameters increases as the sample size increases.

In this chapter, we generalized the methods for 1-way ANOVA in Sun et al. (2012) and give approximations to the Bayes factors under the 2-way ANOVA models with main effects only. Under different cases, these Bayes factors are proved to be consistent with proper choices of hyper-priors, which supports the use of our proposed prior in ANOVA models.

4.2 Model settings

Consider a series of 2-way ANOVA models without the interaction effects $\{M_3^{(N)} : N = 1, 2, \dots\}$ with each model having the form:

$$\begin{aligned} \boldsymbol{y} &= \mathbf{1_{p_1}} \otimes \mathbf{1_{p_2}} \otimes \mathbf{1_k}\mu + \mathbf{I_{p_1}} \otimes \mathbf{1_{p_2}} \otimes \mathbf{1_k}\boldsymbol{\beta_1} + \mathbf{1_{p_1}} \otimes \mathbf{I_{p_2}} \otimes \mathbf{1_k}\boldsymbol{\beta_2} + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \end{aligned} \tag{4.2}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector (let $n = kp_1p_2$), μ is the unknown grand mean parameter, for $i = 1, 2$, $\boldsymbol{\beta}_i$ is a $p_i \times 1$ vector of unknown parameters, and $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$.

We assume that k, p_1, and p_2 are all nondecreasing functions of N satisfying $k \geq 1$, $p_1 \geq 2$, and $p_2 \geq 2$. We also assume that at least one of them goes to infinity as $N \to \infty$, which guarantees that $n \to \infty$ as $N \to \infty$. Let $\lim_{N\to\infty} \frac{p_i}{n} = c_i$ for $i = 1, 2$, where $0 \leq c_i < 1$. We consider two scenarios according to the values of c_1 and c_2. Scenario I is when $c_1 = c_2 = 0$, which means that as $N \to \infty$, either k goes to infinity or both p_1 and p_2 go to infinity or both. Scenario II is when one of c_1 and c_2 is non-zero and the other one is zero, which occurs only when k and one of the p_i's are fixed whereas the other p_i goes to infinity as $N \to \infty$.

For the simplicity of notation, we omit the superscript N unless necessary and denote the full model with M_3 and its sub-models as follows.

$$\begin{aligned}
M_0 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{\epsilon}, \\
M_1 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \\
M_2 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.
\end{aligned}$$

The proposed prior for these models is:

$$\pi(\mu, \sigma^2) = \frac{1}{\sigma^2}, \tag{4.3}$$

$$\boldsymbol{\beta}_i|\sigma^2, g_i \overset{indep}{\sim} N_{p_i}(\boldsymbol{0}, g_i\sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}), \text{ for } i = 1, 2, \tag{4.4}$$

$$g_i \overset{indep.}{\sim} \text{Inv-Gamma}(1/2, nb_i), \text{ for } i = 1, 2. \tag{4.5}$$

For $i = 1, 2$, the Bayes factor between M_i and M_0 is:

$$B_{i0} = \int_0^\infty (1+g_i)^{-\frac{p_i-1}{2}}\Big(1 - \frac{g_i}{1+g_i}\frac{SSR_i}{SST}\Big)^{-\frac{n-1}{2}}\pi(g_i)dg_i, \tag{4.6}$$

and the Bayes factor between M_3 and M_0 is:

$$B_{30} = \int_0^\infty\int_0^\infty \Big(1 - \sum_{i=1}^{2}\frac{g_i}{1+g_i}\frac{SSR_i}{SST}\Big)^{-\frac{n-1}{2}}\prod_{i=1}^{2}[(1+g_i)^{-\frac{p_i-1}{2}}\pi(g_i)]dg_1dg_2, \tag{4.7}$$

where $SST = \boldsymbol{y}'(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{y}$ is the commonly used corrected total sum of squares, and for $i = 1, 2$, $SSR_i = \boldsymbol{y}'[\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' - \frac{1}{n}\boldsymbol{J}_n]\boldsymbol{y}$ is the regression sum of squares corresponding to $\boldsymbol{\beta}_i$. Note that the sums of squares and the Bayes factors all depend on N. In this chapter, we discuss the asymptotic behavior of the Bayes factors as $N \to \infty$.

In the following discussion, we assume that M_3 is the true model, which contains the situations when any sub-model is true. Let the true value of σ^2 be $\tilde{\sigma}^2$ which does not depend on N. For $i = 1, 2$, we assume that either $\boldsymbol{\beta}_i \sim N_{p_i}\big(\boldsymbol{0}, \tilde{\sigma}_i^2\boldsymbol{I}_{p_i}\big)$ where $\tilde{\sigma}_i^2$ is fixed or $\boldsymbol{\beta}_i$ are fixed effects and $\tau_{\boldsymbol{\beta}_i} = \frac{1}{p_i-1}\sum_j(\beta_{ij}-\bar{\beta}_i)^2 \to c_{\boldsymbol{\beta}_i}$ as $N \to \infty$, where $c_{\boldsymbol{\beta}_i}$ is a nonnegative constant. Furthermore, we let $\tilde{g}_i = \tilde{\sigma}_i^2/\tilde{\sigma}^2$ if $\boldsymbol{\beta}_i$ are random effects and $\tilde{g}_i = c_{\boldsymbol{\beta}_i}/\tilde{\sigma}^2$ if $\boldsymbol{\beta}_i$ are fixed effects. Fact 1 summarizes the distributions of the sum of squares that are of interest.

Fact 1. *(a) The sample distribution of SSR_i is*

$$SSR_i \sim \begin{cases} \tilde{\sigma}^2\chi^2_{p_i-1}(\frac{n(p_i-1)}{p_i\tilde{\sigma}^2}\tau_{\boldsymbol{\beta}_i}), & \text{if } \boldsymbol{\beta}_i \text{ are fixed effects,} \\ (\tilde{\sigma}^2+\frac{n}{p_i}\tilde{\sigma}_i^2)\chi^2_{p_i-1}, & \text{if } \boldsymbol{\beta}_i \sim N_{p_i}\big(\mathbf{0},\tilde{\sigma}_i^2\boldsymbol{I}_{p_i}\big). \end{cases}$$

(b) The sample distribution of $SSE^* \equiv SST - SSR_1 - SSR_2$ is

$$SSE^* \sim \tilde{\sigma}^2\chi^2_{n-p_1-p_2+1}.$$

We also need the following lemma in our discussion.

Lemma 5. *For $i = 1$ or 2, consider the following situations.*

(a1)$\boldsymbol{\beta}_i \sim N_{p_i}(\mathbf{0},\tilde{\sigma}_i^2\boldsymbol{I}_{p_i})$ where $\tilde{\sigma}_i^2 = 0$ or $p_i \to \infty$ as $N \to \infty$.

(a2)$\boldsymbol{\beta}_i \sim N_{p_i}(\mathbf{0},\tilde{\sigma}_i^2\boldsymbol{I}_{p_i})$ when $\tilde{\sigma}_i^2 \neq 0$ and p_i is fixed as $N \to \infty$.

(b)$\boldsymbol{\beta}_i$ are fixed effects.

When (a1) or (b) holds, $\frac{SSR_i}{SSE^*}$ converges in probability to a nonnegative constant as $N \to \infty$. When (a2) holds, $\frac{SSR_i}{SSE^*} = O_p(1)$ as $N \to \infty$.

Proof. When $\boldsymbol{\beta}_i \sim N_{p_i}\big(\mathbf{0},\tilde{\sigma}_i^2\boldsymbol{I}_{p_i}\big)$,

$$\frac{SSR_i}{SSE^*} = \Big(1+\frac{n}{p_i}\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}^2}\Big)\frac{\chi^2_{p_i-1}}{\chi^2_{n-p_1-p_2+1}},$$

where $\chi^2_{p_i-1}$ and $\chi^2_{n-p_1-p_2+1}$ are two independent chi-square random variables with the corresponding degrees of freedom. As $N \to \infty$, $n-p_1-p_2+1 = (k-1)p_1p_2 + (p_1-1)(p_2-1) \to \infty$, so $\frac{\chi^2_{n-p_1-p_2+1}}{n-p_1-p_2+1} \xrightarrow{P} 1$. If $p_i \to \infty$ as well, we can obtain that

$\frac{n-p_1-p_2+1}{p_i-1}\frac{\chi^2_{p_i-1}}{\chi^2_{n-p_1-p_2+1}} \xrightarrow{P} 1$ as $N \to \infty$, which leads to

$$\frac{SSR_i}{SSE^*} \xrightarrow{P} \frac{c_i}{1-c_1-c_2} + \frac{1}{1-c_1-c_2}\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}^2}.$$

If p_i is fixed (or goes to a constant) as $N \to \infty$, then $\frac{SSR_i}{SSE^*} \xrightarrow{P} 0$ when $\tilde{\sigma}_i^2 = 0$. If $\tilde{\sigma}_i^2 \neq 0$, then for any $K > 0$,

$$\begin{aligned} & P\Big(\Big|\frac{SSR_i}{SSE^*}\Big| \geq K\Big) \\ \leq \; & P\Big(\Big|\frac{(n-p_1-p_2+1)\tilde{\sigma}^2}{SSE^*}\Big| \geq 2\Big) + P\Big(\Big|\frac{SSR_i}{(n-p_1-p_2+1)\tilde{\sigma}^2}\Big| \geq \frac{K}{2}\Big). \end{aligned}$$

The first component goes to zero as $N \to \infty$ since $\frac{(n-p_1-p_2+1)\tilde{\sigma}^2}{SSE^*} \xrightarrow{P} 1$, whereas for sufficiently large N,

$$\begin{aligned} P\Big(\Big|\frac{SSR_i}{(n-p_1-p_2+1)\tilde{\sigma}^2}\Big| \geq \frac{K}{2}\Big) &= P\Big(\chi^2_{p_i-1} \geq \frac{(1-\frac{p_1}{n}-\frac{p_2}{n}+\frac{1}{n})K}{2(\frac{1}{n}+\frac{1}{p_i}\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}^2})}\Big) \\ &\leq P\Big(\chi^2_{p_i-1} \geq \frac{1-c_1-c_2}{\frac{4}{p_i}\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}^2}}K\Big), \end{aligned}$$

which does not depend on N and goes to zero as $K \to \infty$. Therefore, $\frac{SSR_i}{SSE^*} = O_p(1)$ as $N \to \infty$.

When condition (b) holds, SSE^* satisfies a chi-square distribution and SSR_i satisfies a non-central chi-square distribution. SSR_i can be written as

$$\begin{aligned} \frac{SSR_i}{\tilde{\sigma}^2} &= \Big(Z_1 + \sqrt{\frac{\tau_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}\frac{p_i-1}{p_i}n}\Big)^2 + Z_2^2 + \cdots + Z_{p_i-1}^2 \\ &= Z_1^2 + \cdots + Z_{p_i-1}^2 + n\frac{\tau_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}\frac{p_i-1}{p_i} + 2\sqrt{\frac{\tau_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}\frac{p_i-1}{p_i}n}Z_1, \end{aligned}$$

where $Z_1, \ldots, Z_{p_i-1} \overset{i.i.d.}{\sim} N(0,1)$. By the weak law of large numbers, as $N \to \infty$,

$$\frac{Z_1^2 + \cdots + Z_{p_i-1}^2}{n} \xrightarrow{P} c_i.$$

For the other two terms, as $N \to \infty$,

$$2\sqrt{\frac{\tau_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}\frac{p_i - 1}{p_i}}\frac{Z_1}{\sqrt{n}} \overset{P}{\to} 0,$$

$$\frac{p_i - 1}{p_i}\frac{\tau_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2} \to \begin{cases} \frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}, & \text{if } p_i \to \infty \text{ as } N \to \infty, \\ \frac{p_i - 1}{p_i}\frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}, & \text{if } p_i \text{ goes to a constant as } N \to \infty. \end{cases}$$

Similar to before, as $N \to \infty$,

$$\frac{SSE^*}{n\tilde{\sigma}^2} \overset{P}{\to} 1 - c_1 - c_2.$$

Hence,

$$\frac{SSR_i}{SSE^*} \overset{P}{\to} \begin{cases} \frac{1}{1-c_1-c_2}\left(\frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2} + c_i\right), & \text{if } p_i \to \infty \text{ as } N \to \infty, \\ \frac{1}{1-c_1-c_2}\left(\frac{p_i-1}{p_i}\frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2} + c_i\right), & \text{if } p_i \text{ goes to a constant as } N \to \infty. \end{cases}$$

□

4.3 Scenario I

4.3.1 Gamma-type approximation

Under Scenario I, we use the Gamma-type approximation similar to Sun et al. (2012) as proved in the following lemma.

Lemma 6. *Suppose $Y_N \sim \Gamma(\frac{m}{2}, 1), N = 1, 2, \ldots$*

(a) For fixed $m \geq 2$, if $M_N \to \infty$ as $N \to \infty$, then for any $\epsilon > 0$, $P(Y_N \geq M_N\epsilon) \to 0$.

(b) If as $N \to \infty$, $m = m_N \to \infty$, $M_N \to \infty$ and $m_N/M_N \to 0$, then for any $\epsilon > 0$, $P(Y_N \geq M_N\epsilon) \to 0$.

Proof. Part (a) is obvious. For Part (b), when N is large enough,

$$\begin{aligned}
P\left(Y_N \geq M\epsilon\right) &= P\left(Y_N - \frac{m}{2} \geq M\epsilon - \frac{m}{2}\right) \\
&\leq P\left(|Y_N - \frac{m}{2}| \geq |M\epsilon - \frac{m}{2}|\right) \\
&\leq \frac{E\left(Y_N - \frac{m}{2}\right)^2}{\left(M\epsilon - \frac{m}{2}\right)^2} \\
&= \frac{m}{2M^2}\frac{1}{\left(\epsilon - \frac{m}{2M}\right)^2} \\
&\to 0, \text{ as } N \to \infty.
\end{aligned}$$

$\square$

Then we can obtain approximations of B_{i0} $(i = 1, 2, 3)$.

Proposition 1. *(a) Define*

$$\begin{aligned}
h(w_1, w_2) &= \frac{1}{2}\log\left(1 - \sum_{i=1}^{2}\frac{1}{1+w_i}\frac{SSR_i}{SST}\right) + \sum_{i=1}^{2} b_i w_i + \frac{1}{2n}\sum_{i=1}^{2}(p_i - 1)\log(1 + w_i), \\
T(w_1, w_2) &= \sqrt{1 - \sum_{i=1}^{2}\frac{1}{1+w_i}\frac{SSR_i}{SST}}.
\end{aligned}$$

Then,

$$B_{30} = n\sqrt{\frac{b_1 b_2}{\pi^2}}\int_0^\infty\int_0^\infty w_1^{\frac{p_1}{2}-1} w_2^{\frac{p_2}{2}-1}\exp\{-nh(w_1, w_2)\}T(w_1, w_2)dw_1 dw_2.$$

(b) Suppose that $p_1 = o(nb_1)$ as $N \to \infty$. Then $\forall \epsilon > 0$,

$$\frac{F_1(w_2)}{F_2(w_2)} \equiv \frac{\int_\epsilon^\infty w_1^{\frac{p_1}{2}-1}\frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}}dw_1}{\int_0^\epsilon w_1^{\frac{p_1}{2}-1}\frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}}dw_1} \to 0 \tag{4.8}$$

uniformly w.r.t. w_2 as $N \to \infty$. Symmetrically, suppose that $p_2 = o(nb_2)$ as $N \to \infty$, *then* $\forall \epsilon > 0$,

$$\frac{\int_{\epsilon}^{\infty} w_2^{\frac{p_2}{2}-1} \frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}} dw_2}{\int_0^{\epsilon} w_2^{\frac{p_2}{2}-1} \frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}} dw_2} \to 0 \tag{4.9}$$

uniformly w.r.t. w_1 *as* $N \to \infty$.

(c) Suppose that for $i = 1, 2$, $p_i = o(nb_i)$ *as* $N \to \infty$. *Define*

$$R_{30} = n^{-\frac{p_1+p_2-2}{2}} \Gamma(\frac{p_1}{2})\Gamma(\frac{p_2}{2})\sqrt{\frac{b_1 b_2}{\pi^2}}\Big(\frac{SSE^*}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{2}\Big(b_i + \frac{1}{2}\frac{SSR_i}{SSE^*} + \frac{p_i - 1}{2n}\Big)^{-\frac{p_i}{2}}.$$

Then,

$$\frac{1}{max(p_1, p_2)} \log\Big(\frac{B_{30}}{R_{30}}\Big) \xrightarrow{P} 0.$$

(d) For $i = 1, 2$, *define*

$$R_{i0} = n^{-\frac{p_i-1}{2}} \Gamma(\frac{p_i}{2})\sqrt{\frac{b_i}{\pi}}\Big(\frac{SST - SSR_i}{SST}\Big)^{-\frac{n-1}{2}} \Big(b_i + \frac{1}{2}\frac{SSR_i}{SST - SSR_i} + \frac{p_i - 1}{2n}\Big)^{-\frac{p_i}{2}}.$$

Suppose for $i = 1$ *or* 2, $p_i = o(nb_i)$ *as* $N \to \infty$. *Then,*

$$\frac{1}{p_i} \log\Big(\frac{B_{i0}}{R_{i0}}\Big) \xrightarrow{P} 0.$$

Proof. (a) Consider the transformation $w_i = 1/g_i$ $(i = 1, 2)$ in (4.7), then

$$\begin{aligned}
B_{30} &= \frac{1}{n^2}\int_0^{\infty}\int_0^{\infty}\Big(1 - \sum_{i=1}^{2}\frac{g_i}{1+g_i}\frac{SSR_i}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{2}\frac{\mathrm{IG}(\frac{g_i}{n}; \frac{1}{2}, b_i)}{(1+g_i)^{\frac{p_i-1}{2}}} dg_1 dg_2 \\
&= \frac{1}{n^2}\int_0^{\infty}\int_0^{\infty}\Big(1 - \sum_{i=1}^{2}\frac{1}{1+w_i}\frac{SSR_i}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{2}\frac{\mathrm{IG}(\frac{1}{nw_i}; \frac{1}{2}, b_i) w_i^{\frac{p_i-5}{2}}}{(1+w_i)^{\frac{p_i-1}{2}}} dw_1 dw_2 \\
&= n\sqrt{\frac{b_1 b_2}{\pi^2}}\int_0^{\infty}\int_0^{\infty} w_1^{\frac{p_1}{2}-1} w_2^{\frac{p_2}{2}-1}\exp\{-nh(w_1, w_2)\} T(w_1, w_2) dw_1 dw_2,
\end{aligned}$$

(b) When $w_1 \in (0, \epsilon)$, by Taylor expansion,

$$\begin{aligned} T(w_1, w_2) &\geq T(0, w_2), \\ h(w_1, w_2) &= h(0, w_2) + h_1'(\xi, w_2)w_1 \leq h(0, w_2) + h_1'(0, w_2)w_1, \end{aligned}$$

where $\xi \in (0, \epsilon)$ and for $i = 1, 2$,

$$\begin{aligned} h_i'(w_1, w_2) &\equiv \frac{\partial h(w_1, w_2)}{\partial w_i} \\ &= b_i + \frac{1}{2} \frac{1}{\left(1 - \sum_{j=1}^{2} \frac{1}{1+w_j} \frac{SSR_j}{SST}\right)} \frac{1}{(1+w_i)^2} \frac{SSR_i}{SST} + \frac{p_i - 1}{2n} \frac{1}{1+w_i}. \end{aligned}$$

Hence,

$$F_2(w_2) \geq \frac{T(0, w_2)}{\exp[nh(0, w_2)]} \int_0^{\epsilon} \frac{w_1^{\frac{p_1}{2}-1}}{\exp[nh_1'(0, w_2)w_1]} dw_1.$$

Define

$$H_2(w_2) = T(0, w_2) \exp[-nh(0, w_2)] \frac{\Gamma(\frac{p_1}{2})}{[nh_1'(0, w_2)]^{p_1/2}},$$

then

$$\begin{aligned} \frac{F_2(w_2)}{H_2(w_2)} &\geq \frac{[nh_1'(0, w_2)]^{p_1/2}}{\Gamma(\frac{p_1}{2})} \int_0^{\epsilon} \frac{w_1^{\frac{p_1}{2}-1}}{\exp[nh_1'(0, w_2)w_1]} dw_1 \\ &\geq \int_0^{\epsilon} \frac{(nb_1)^{p_1/2}}{\Gamma(\frac{p_1}{2})} w_1^{\frac{p_1}{2}-1} \exp[-nb_1 w_1] dw_1 \to 1, \end{aligned}$$

as $N \to \infty$ according to Lemma 6. Therefore, it suffices to show that as $N \to \infty$,

$$\frac{F_1(w_2)}{H_2(w_2)} = \frac{\exp[nh(0, w_2)]}{T(0, w_2)} \frac{[nh_1'(0, w_2)]^{\frac{p_1}{2}}}{\Gamma(\frac{p_1}{2})} \int_{\epsilon}^{\infty} w_1^{\frac{p_1}{2}-1} \frac{T(w_1, w_2)}{\exp\{nh(w_1, w_2)\}} dw_1 \to 0.$$

When $w_1 \in (\epsilon, \infty)$,

$$\begin{aligned}
& \frac{T(w_1,w_2)}{T(0,w_2)} \exp\{-n[h(w_1,w_2)-h(0,w_2)]\} \\
\leq \ & \Big(1+\big(1-\frac{1}{1+\epsilon}\big)\frac{SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big)^{-\frac{n-1}{2}} \exp(-nb_1w_1)(1+w_1)^{-\frac{p_1-1}{2}}.
\end{aligned}$$

Since

$$\int_\epsilon^\infty w_1^{\frac{p_1}{2}-1}(1+w_1)^{-\frac{p_1-1}{2}} \exp(-nb_1w_1)dw_1 \leq \frac{\Gamma\big(\frac{p_1}{2}\big)}{\big(\frac{nb_1}{2}\big)^{\frac{p_1}{2}}} \exp\big(-\frac{nb_2\epsilon}{2}\big),$$

we only need to show that as $N \to \infty$,

$$\exp\big(-\frac{nb_2\epsilon}{2}\big)\Big(1+\frac{\frac{\epsilon}{1+\epsilon}SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big)^{-\frac{n-1}{2}}\Big[\frac{2h_1'(0,w_2)}{b_1}\Big]^{\frac{p_1}{2}} \to 0.$$

Note that

$$\begin{aligned}
\frac{2h_1'(0,w_2)}{b_1} &= 2+\frac{1}{b_1}\frac{SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}+\frac{p_1-1}{nb_1} \\
&\leq \big(1+1+\frac{p_1-1}{nb_1}\big)\Big(1+\frac{1}{b_1}\frac{SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big).
\end{aligned}$$

As $N \to \infty$, because $p_1 = o(nb_1)$,

$$\exp\big(-\frac{nb_1\epsilon}{2}\big)\big(1+1+\frac{p_1-1}{nb_1}\big)^{\frac{p_1}{2}} \to 0.$$

Similarly, when N is large enough such that $\frac{\epsilon}{1+\epsilon} > \frac{p_1}{b_1(n-1)}$,

$$\begin{aligned}
& \frac{2}{n-1}\log\Big[\Big(1+\frac{1}{b_1}\frac{SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big)^{\frac{p_1}{2}}\Big] \\
\leq \ & \log\Big(1+\frac{p_1}{b_1(n-1)}\frac{SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big) \\
\leq \ & \log\Big(1+\frac{\frac{\epsilon}{1+\epsilon}SSR_1}{SST-SSR_1-\frac{1}{1+w_2}SSR_2}\Big).
\end{aligned}$$

Therefore, (4.8) is proved and (4.9) can be proved similarly.

(c) For any $\epsilon > 0$, consider

$$\begin{aligned} & J_1 + J_2 + J_3 + J_4 \\ = & \Big(\int_0^\epsilon \int_0^\epsilon + \int_0^\epsilon \int_\epsilon^\infty + \int_\epsilon^\infty \int_0^\epsilon + \int_\epsilon^\infty \int_\epsilon^\infty\Big) \sqrt{\frac{b_1 b_2}{\pi^2}} \frac{nT(w_1, w_2)}{\exp\{nh(w_1, w_2)\}} \prod_{i=1}^2 w_i^{\frac{p_i}{2}-1} dw_i. \end{aligned}$$

From (4.8) and (4.9), as $N \to \infty$,

$$\frac{J_2}{J_1}, \frac{J_3}{J_1}, \frac{J_4}{J_1} \to 0,$$

Therefore,

$$\log\Big(\frac{J_1}{B_{30}}\Big) \to 0. \tag{4.10}$$

When $w_1, w_2 \in (0, \epsilon)$, by Taylor expansion,

$$h(w_1, w_2) = h(0,0) + h_1'(\xi_1, \xi_2) w_1 + h_2'(\xi_1, \xi_2) w_2,$$

where $\xi_1, \xi_2 \in (0, \epsilon)$, so

$$h(0,0) + h_1'(\epsilon, \epsilon) w_1 + h_2'(\epsilon, \epsilon) w_2 \le h(w_1, w_2) \le h(0,0) + h_1'(0,0) w_1 + h_2'(0,0) w_2.$$

Moreover, when $w_1, w_2 \in (0, \epsilon)$,

$$T(\epsilon, \epsilon) \ge T(w_1, w_2) \ge T(0,0).$$

Therefore,

$$\prod_{i=1}^2 \frac{[nh_i'(0,0)]^{\frac{p_i}{2}}}{\Gamma(\frac{p_i}{2})} \int_0^\epsilon \frac{w_i^{\frac{p_i}{2}-1}}{\exp[nh_i'(0,0) w_i]} dw_i \le \frac{J_1}{R_{30}}, \tag{4.11}$$

and on the other side,

$$\begin{aligned}
\frac{J_1}{R_{30}} &\le \frac{T(\epsilon,\epsilon)}{T(0,0)} \prod_{i=1}^{2} \frac{[nh_i'(0,0)]^{\frac{p_i}{2}}}{\Gamma(\frac{p_i}{2})} \int_0^{\epsilon} \frac{w_i^{\frac{p_i}{2}-1}}{\exp[nh_i'(\epsilon,\epsilon)w_i]} dw_i \\
&\le \frac{T(\epsilon,\epsilon)}{T(0,0)} \prod_{i=1}^{2} \Big[\frac{h_i'(0,0)}{h_i'(\epsilon,\epsilon)}\Big]^{\frac{p_i}{2}}.
\end{aligned} \tag{4.12}$$

Using Lemma 6, it can be shown that for $i = 1, 2$,

$$\lim_{N\to\infty} \frac{[nh_i'(0,0)]^{\frac{p_i}{2}}}{\Gamma(\frac{p_i}{2})} \int_0^{\epsilon} \frac{w_i^{\frac{p_i}{2}-1}}{\exp[nh_i'(0,0)w_i]} dw_i = 1,$$

so as $N \to \infty$,

$$\log\Big(\frac{J_1}{R_{30}}\Big) \ge -o(1). \tag{4.13}$$

From (4.12), note that

$$\frac{T(\epsilon,\epsilon)}{T(0,0)} = \sqrt{1 + \frac{\epsilon}{1+\epsilon}\frac{SSR_1 + SSR_2}{SSE^*}}.$$

Moreover, for $i = 1, 2$,

$$\begin{aligned}
\frac{h_i'(0,0)}{h_i'(\epsilon,\epsilon)} &= 1 + \frac{\frac{SSR_i}{2}\big(\frac{1}{SSE^*} - \frac{1}{(1+\epsilon)(SSE^*+\epsilon SST)}\big) + \frac{(p_i-1)\epsilon}{2n(1+\epsilon)}}{b_i + \frac{SSR_i}{2(1+\epsilon)(SSE^*+\epsilon SST)} + \frac{p_i-1}{2n(1+\epsilon)}} \\
&\le (1+\epsilon)^2\Big(1 + \frac{\epsilon}{1+\epsilon}\frac{SSR_1 + SSR_2}{SSE^*}\Big).
\end{aligned}$$

Hence,

$$\frac{2}{p_1 + p_2 + 1}\log\Big(\frac{J_1}{R_{30}}\Big) \le 2\epsilon + \frac{\epsilon}{1+\epsilon}\frac{SSR_1 + SSR_2}{SSE^*}. \tag{4.14}$$

Then, (4.13) and (4.10) suggest that

$$\log\left(\frac{B_{30}}{R_{30}}\right) \geq -o(1). \tag{4.15}$$

The right hand side of (4.15) goes to zero as $N \to \infty$. (4.14) and (4.10) suggest that

$$\frac{2}{p_1+p_2+1}\log\left(\frac{B_{30}}{R_{30}}\right) \;\leq\; 2\epsilon + \frac{\epsilon}{1+\epsilon}\frac{SSR_1+SSR_2}{SSE^*} + o(1). \tag{4.16}$$

When $\frac{SSR_1+SSR_2}{SSE^*}$ converges in probability to some constant, say c, the right hand side of (4.16) converges in probability to $2\epsilon + \frac{\epsilon}{1+\epsilon}c$. Since ϵ can be chosen arbitrarily small, $\frac{2}{p_1+p_2+1}\log\left(\frac{B_{30}}{R_{30}}\right)$ converges in probability to zero. When $\frac{SSR_1+SSR_2}{SSE^*} = O_p(1)$, $\forall \delta, d > 0$,

$$\begin{aligned}
&P\Big(\frac{2}{p_1+p_2+1}\log\left(\frac{B_{30}}{R_{30}}\right) \geq d\Big) \\
\leq\;& P\Big(2\epsilon + \epsilon\frac{SSR_1+SSR_2}{SSE^*} + o(1) \geq d\Big) \\
\leq\;& P\Big(\epsilon\Big(2 + \frac{SSR_1+SSR_2}{SSE^*}\Big) \geq d/2\Big) + P(o(1) \geq d/2).
\end{aligned}$$

Since the inequality above holds for any $\epsilon > 0$, we can first choose sufficiently small ϵ such that for some integer N_1, the first component is less than δ for any $N \geq N_1$. Then for this ϵ, there is some integer N_2 such that the second component is zero for any $N \geq N_2$. Then for $N \geq \max(N_1, N_2)$, $P\Big(\frac{2}{p_1+p_2+1}\log\left(\frac{B_{30}}{R_{30}}\right) \geq d\Big) < \delta$. Therefore, $\frac{2}{p_1+p_2+1}\log\left(\frac{B_{30}}{R_{30}}\right)$ converges to zero in probability. Part (c) is proved, and Part (d) can be proved using similar procedures. □

4.3.2 Consistency of the Bayes factors

In this part, we prove the consistency of the Bayes factors under Scenario I. We consider the random effects model first.

Theorem 6. *For $i = 1$ or 2, assume that $p_i = o(nb_i)$, $b_i = O(1)$ as $N \to \infty$, and suppose that $\boldsymbol{\beta}_i$ are random effects.*

(a) When p_i is fixed as $N \to \infty$, under M_0,

$$\frac{2}{\log(nb_i)(p_i - 1)} \log B_{i0} \xrightarrow{P} -1, \tag{4.17}$$

and under M_i,

$$\frac{2}{n} \log B_{i0} \xrightarrow{L} \log\left(1 + \frac{\tilde{g}_i}{p_i}\chi^2_{p_i-1}\right). \tag{4.18}$$

(b) When $p_i \to \infty$ as $N \to \infty$, under M_0,

$$\frac{2}{\log(2nb_i/p_i)(p_i - 1)} \log B_{i0} \xrightarrow{P} -1, \tag{4.19}$$

and under M_i,

$$\frac{2}{n} \log B_{i0} \xrightarrow{P} \log(1 + \tilde{g}_i). \tag{4.20}$$

(c) B_{i0} is consistent under both M_i and M_0.

Proof. Without loss of generality, we only consider B_{10}. According to Proposition 1, we only need to prove the results for R_{10} instead of B_{10}. Under M_0 and M_1, $\tilde{g}_2 = 0$, according to Fact 1,

$$\frac{SSR_1}{SST - SSR_1} = \begin{cases} \frac{(1+\frac{n}{p_1}\tilde{g}_1)\chi^2_{p_1-1}}{\chi^2_{p_1(p_2k-1)}} & \text{under } M_1, \\ \frac{\chi^2_{p_1-1}}{\chi^2_{p_1(p_2k-1)}} & \text{under } M_0, \end{cases}$$

where in both cases, $\chi^2_{p_1(p_2k-1)}$ and $\chi^2_{p_1-1}$ are two independent χ^2 random variables with corresponding degrees of freedom.

(a)If p_1 is fixed, under M_0, as $N \to \infty$,

$$\frac{1}{b_1}\left[\frac{SSR_1}{2(SST - SSR_1)} + \frac{p_1 - 1}{2n}\right] \xrightarrow{P} 0, \tag{4.21}$$

$$\frac{1}{SSR_1}\log\left[\left(1 + \frac{SSR_1}{SST - SSR_1}\right)^{\frac{n-1}{2}}\right] \xrightarrow{P} \frac{1}{2}, \tag{4.22}$$

$$\frac{SSR_1}{\log(nb_1)} \xrightarrow{P} 0. \tag{4.23}$$

Therefore, the leading term in $\log(R_{10})$ is $-\frac{p_1-1}{2}\log(nb_1)$, so (4.17) is proved.

Under M_1, since $b_1 = O(1)$, the leading term in $\log(B_{10})$ is $\frac{n-1}{2}\log\left(1 + \frac{SSR_1}{SST-SSR_1}\right)$. (4.18) follows by noticing that $\frac{\chi^2_{p_1(p_2k-1)}}{n} \xrightarrow{P} 1$.

(b) If $p_1 \to \infty$, by Stirling's formula,

$$\frac{\Gamma(\frac{p_1}{2})}{\sqrt{2\pi}(\frac{p_1}{2})^{\frac{p_1-1}{2}}e^{-\frac{p_1}{2}}} \to 1.$$

Also,

$$\frac{SSR_1}{SST - SSR_1} \xrightarrow{P} \tilde{g}_1 \text{ under } M_1,$$

$$\frac{n - p_1}{p_1 - 1}\frac{SSR_1}{SST - SSR_1} \xrightarrow{P} 1 \text{ under } M_0.$$

Therefore, under M_0,

$$\frac{n-1}{p_1-1}\log\left(1 + \frac{SSR_1}{SST - SSR_1}\right) \xrightarrow{P} 1.$$

Arguments similar to Part (a) carry out, after taking $\Gamma(\frac{p_1}{2})$ into account, the leading term in $\log(B_{10})$ is now $-\frac{p_1-1}{2}\log(2nb_1/p_1)$. Under M_1, since $\frac{p_1}{n}\log(\frac{n}{p_1}) \to 0$, $\frac{n-1}{2}\log\left(1 + \frac{SSR_1}{SST-SSR_1}\right)$ is still the leading term in $\log(B_{10})$. Hence, (4.19) and (4.20) are proved.

The consistency properties in Part (c) are immediate consequences of Parts (a) and (b). $\square$

Now consider B_{i0} $(i = 1, 2)$ under fixed effect models.

Theorem 7. *For $i = 1$ or 2, assume that $p_i = o(nb_i)$, $b_i = O(1)$ as $N \to \infty$, and suppose that $\boldsymbol{\beta}_i$ are fixed effects.*

(a) When p_i is fixed as $N \to \infty$, under M_0,

$$\frac{2}{\log(nb_i)(p_i - 1)} \log B_{i0} \xrightarrow{P} -1, \tag{4.24}$$

and under M_i,

$$\frac{2}{n} \log B_{i0} \xrightarrow{P} \log\left(1 + \frac{p_i - 1}{p_i} \tilde{g}_i\right). \tag{4.25}$$

(b) When $p_i \to \infty$ as $N \to \infty$, under M_0,

$$\frac{2}{\log(2nb_i/p_i)(p_i - 1)} \log B_{i0} \xrightarrow{P} -1, \tag{4.26}$$

and under M_i,

$$\frac{2}{n} \log B_{i0} \xrightarrow{P} \log(1 + \tilde{g}_i). \tag{4.27}$$

(c) B_{i0} is consistent under both M_i and M_0.

Proof. Under M_0, the distribution of SSR_i is the same as it is under random effect models, so the asymptotic properties are the same as well. Under M_i, with steps similar to those in Lemma 5, we know that

$$\frac{SSR_i}{SST - SSR_i} \xrightarrow{P} \begin{cases} \frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}, & \text{if } p_i \to \infty \text{ as } N \to \infty, \\ \frac{p_i - 1}{p_i} \frac{c_{\boldsymbol{\beta}_i}}{\tilde{\sigma}^2}, & \text{if } p_i \text{ is fixed as } N \to \infty. \end{cases}$$

Therefore, the results can be proved similarly. $\square$

With similar arguments, we can also prove the consistency of B_{30}.

Theorem 8. *For $i = 1, 2$, assume that $p_i = o(nb_i)$ and $b_i = O(1)$ as $N \to \infty$.*

(a) Under M_0, as $N \to \infty$,

$$\frac{2}{A_1 + A_2} \log B_{30} \xrightarrow{P} -1, \tag{4.28}$$

where for $i = 1, 2$, $A_i = (p_i - 1)\log(nb_i)$ if p_i is fixed as $N \to \infty$; $A_i = (p_i - 1)\log(2nb_i/p_i)$ if $p_i \to \infty$ as $N \to \infty$.

(b) Under M_3, as $N \to \infty$,

$$\frac{2}{n} \log B_{30} \xrightarrow{L} \log(1 + C_1 + C_2), \tag{4.29}$$

where for $i = 1, 2$, if p_i is fixed as $N \to \infty$, $C_i = \frac{\tilde{g}_i}{p_i}\chi^2_{p_i - 1}$ when $\boldsymbol{\beta}_i$ are random effects and $C_i = \frac{p_i - 1}{p_i}\tilde{g}_i$ when $\boldsymbol{\beta}_i$ are fixed effects; if $p_i \to \infty$ as $N \to \infty$, $C_i = \tilde{g}_i$.

(c) B_{30} is consistent under both M_0 and M_3.

Theorem 9. *(a) Assume that $p_1 = o(nb_1)$ and $b_1 = O(1)$ as $N \to \infty$. Under M_2, when p_1 is fixed as $N \to \infty$,*

$$\frac{2}{\log(nb_1)(p_1 - 1)} \log B_{10} \xrightarrow{P} -1, \tag{4.30}$$

and when $p_1 \to \infty$ as $N \to \infty$,

$$\frac{2}{\log(2nb_1/p_1)(p_1 - 1)} \log B_{10} \xrightarrow{P} -1. \tag{4.31}$$

(b) Assume that for $i = 1, 2$, $p_i = o(nb_i)$ and $b_i = O(1)$ as $N \to \infty$. Under M_2, when p_2 is fixed as $N \to \infty$,

$$\frac{2}{n} \log B_{21} \begin{cases} \xrightarrow{L} \log\left(1 + \frac{\tilde{g}_2}{p_2}\chi^2_{p_2 - 1}\right), & \text{if } \boldsymbol{\beta}_2 \text{ are random effects,} \\ \xrightarrow{P} \log\left(1 + \frac{p_2 - 1}{p_2}\tilde{g}_2\right), & \text{if } \boldsymbol{\beta}_2 \text{ are fixed effects,} \end{cases}$$

and when $p_2 \to \infty$ *as* $N \to \infty$,

$$\frac{2}{n} \log B_{21} \xrightarrow{P} \log(1 + \tilde{g}_2).$$

(c) Under the same conditions as in Part (b), similar results hold for B_{12} *under* M_1.

(d) B_{21} *is consistent under both* M_1 *and* M_2.

Proof. (a) Under M_2, if $\boldsymbol{\beta}_1$ are random effects,

$$\frac{SSR_1}{SST - SSR_1} = \frac{\chi^2_{p_1-1}}{\chi^2_{n-p_1-p_2+1} + (1 + \frac{n}{p_2}\tilde{g}_2)\chi^2_{p_2-1}}. \tag{4.32}$$

If $\boldsymbol{\beta}_1$ are fixed effects,

$$\frac{SSR_1}{SST - SSR_1} = \frac{\chi^2_{p_1-1}}{\chi^2_{n-p_1-p_2+1} + \chi^2_{p_2-1}(\frac{n(p_2-1)}{p_2\tilde{\sigma}^2}\tau_{\boldsymbol{\beta}_2})}. \tag{4.33}$$

In both cases, the denominator is greater than a chi-square random variable with $(n-p_1)$ degrees of freedom, so $\left(\frac{SST-SSR_1}{SST}\right)^{-\frac{n-1}{2}}$ is still a lower order term in $\log(B_{10})$. Therefore, when p_1 is fixed, $-\frac{p_1-1}{2}\log(nb_1)$ is the leading term in $\log(B_{10})$, whereas when $p_1 \to \infty$ as $N \to \infty$, $-\frac{p_1-1}{2}\log(2nb_1/p_1)$ is the leading term. Hence, (4.30) and (4.31) holds.

To prove Part (b), just note that from Part (a), under M_2

$$\frac{1}{n} \log B_{10} \xrightarrow{P} 0,$$

as $N \to \infty$. The results follow from Theorems 6 and 7.

With similar steps, we can prove Part (c), and Part (d) follows from Parts (b) and (c). $\square$

Theorem 10. *For* $i = 1, 2$, *assume that* $p_i = o(nb_i)$ *and* $b_i = O(1)$ *as* $N \to \infty$.

(a) Under M_1, when p_2 is fixed as $N \to \infty$,

$$\frac{2}{\log(nb_2)(p_2-1)} \log B_{31} \xrightarrow{P} -1, \tag{4.34}$$

and when $p_2 \to \infty$ as $N \to \infty$,

$$\frac{2}{\log(2nb_2/p_2)(p_2-1)} \log B_{31} \xrightarrow{P} -1. \tag{4.35}$$

(b) Under M_3, if $\boldsymbol{\beta}_2$ are random effects, when p_2 is fixed as $N \to \infty$,

$$\frac{2}{n} \log B_{31} \xrightarrow{L} \log\left(1 + \frac{\tilde{g}_2}{p_2}\chi^2_{p_2-1}\right), \tag{4.36}$$

and when $p_2 \to \infty$ as $N \to \infty$,

$$\frac{2}{n} \log B_{31} \xrightarrow{P} \log(1 + \tilde{g}_2). \tag{4.37}$$

If $\boldsymbol{\beta}_2$ are fixed effects, when p_2 is fixed as $N \to \infty$,

$$\frac{2}{n} \log B_{31} \xrightarrow{P} \log\left(1 + \frac{p_2 - 1}{p_2}\tilde{g}_2\right), \tag{4.38}$$

and when $p_2 \to \infty$ as $N \to \infty$,

$$\frac{2}{n} \log B_{31} \xrightarrow{P} \log(1 + \tilde{g}_2). \tag{4.39}$$

(c) Similar results hold for B_{32}.

(d) B_{31} is consistent under both M_1 and M_3. B_{32} is consistent under both M_2 and M_3.

Proof. (a)

$$B_{30} = n\sqrt{\frac{b_1 b_2}{\pi^2}} \int_0^\infty \int_0^\infty \Big(1 - \sum_{i=1}^{2} \frac{1}{1+w_i} \frac{SSR_i}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{2} \frac{w_i^{\frac{p_i}{2}-1}}{(1+w_i)^{\frac{p_i-1}{2}}} \exp(-nb_i w_i) dw_i$$

where

$$1 - \frac{1}{1+w_2}\frac{SSR_2}{SST} \geq \frac{1 - \sum_{i=1}^{2} \frac{1}{1+w_i}\frac{SSR_i}{SST}}{1 - \frac{1}{1+w_1}\frac{SSR_1}{SST}} \geq 1 - \frac{1}{1+w_2}\frac{SSR_2}{SST - SSR_1}.$$

Denote

$$B'_{20} = \sqrt{\frac{nb_2}{\pi}} \int_0^\infty \frac{w_2^{\frac{p_2}{2}-1}}{(1+w_2)^{\frac{p_2-1}{2}}} \exp(-nb_2 w_2)\Big(1 - \frac{1}{1+w_2}\frac{SSR_2}{SST - SSR_1}\Big)^{-\frac{n-1}{2}} dw_2,$$

then

$$B_{20} \leq B_{31} \leq B'_{20}$$

Define

$$R'_{20} = n^{-\frac{p_2-1}{2}} \Gamma(\frac{p_2}{2}) \sqrt{\frac{b_2}{\pi}} \Big(1 + \frac{SSR_2}{SSE^*}\Big)^{\frac{n-1}{2}} \Big(b_2 + \frac{SSR_2}{2SSE^*} + \frac{p_2 - 1}{2n}\Big)^{-\frac{p_2}{2}}.$$

We can prove that

$$\frac{1}{p_2} \log\Big(\frac{B'_{20}}{R'_{20}}\Big) \xrightarrow{P} 0.$$

Under M_1, the two equations in (a) hold for both B'_{20} and B_{20}, so they also hold for B_{31}.

(b) Since $\frac{1}{n}\log\big(\frac{B_{i0}}{R_{i0}}\big) \xrightarrow{P} 0$ for $i = 1, 2, 3$. We only need to show the results for

$R_{31} = R_{30}/R_{10}$. It can be easily seen that

$$\begin{aligned} R_{31} &= n^{-\frac{p_2-1}{2}}\Gamma(\frac{p_2}{2})\sqrt{\frac{b_2}{\pi}}\Big(1+\frac{SSR_2}{SSE^*}\Big)^{\frac{n-1}{2}}\Big(b_2+\frac{SSR_2}{2SSE^*}+\frac{p_2-1}{2n}\Big)^{-\frac{p_2}{2}} \\ &\quad \left(\frac{b_1+\frac{SSR_1}{SSE^*}+\frac{p_1-1}{2n}}{b_1+\frac{SSR_1}{SST-SSR_1}+\frac{p_1-1}{2n}}\right)^{-\frac{p_1}{2}}, \end{aligned} \tag{4.40}$$

where

$$\Big(1+\frac{SSR_2}{SSE^*}\Big)^{-\frac{p_1}{2}} \le \left(\frac{b_1+\frac{SSR_1}{SSE^*}+\frac{p_1-1}{2n}}{b_1+\frac{SSR_1}{SST-SSR_1}+\frac{p_1-1}{2n}}\right)^{-\frac{p_1}{2}} \le 1.$$

Therefore, under M_3, $\frac{n-1}{2}\log\big(1+\frac{SSR_2}{SSE^*}\big)$ is the leading term in $\log(B_{31})$, so $\log(B_{31})$ behaves asymptotically the same as $\log(B_{20})$ under M_2, which proves the results. Part (c) an be proved similarly. Parts (a), (b), and (c) lead to Part (d). □

4.4 Scenario II

4.4.1 Approximation of the Bayes factors

Next, we discuss Scenario II, which can occur only when $p_1 \to \infty$, p_2 and k are fixed, or when $p_2 \to \infty$, p_1 and k are fixed. Without loss of generality, we consider the Bayes factor under the first situation. We further assume that $p_1 = N$ since n can only increase when p_1 increases now. In this situation, Sun et al. (2012) proved that in order for B_{10} to be consistent, b_1 should satisfy $b_1 \to 0$ as $N \to \infty$. Therefore, we limit the discussion under this condition only. For b_2, we again assume it satisfies $p_2 = o(b_2 n)$. In other words, $b_2 n \to \infty$ as $N \to \infty$. Similar to the theorem for 1-way ANOVA in Sun et al. (2012), we have the following approximation for B_{10}.

Proposition 2. *Assume that $b_1 \to 0$ as $N \to \infty$.*

(a) Let

$$R_{10,1} = \exp\left\{\frac{p_1}{2}\log\left[\frac{(1+\tilde{g}_1)^{p_2k}}{1+p_2k\tilde{g}_1}\right] - \frac{1}{2}\log p_1\right\}\left[\frac{4b_1(1+p_2k\tilde{g}_1)^3}{p_2k(p_2k-1)\tilde{g}_1^3(1+\tilde{g}_1)}\right]^{\frac{1}{2}},$$

then under M_1*,*

$$\log\left(\frac{B_{10}}{R_{10,1}}\right) \xrightarrow{a.s.} 0.$$

(b) Further assume that $p_1^3 b_1^2 \to \infty$ *as* $N \to \infty$*. Let*

$$R_{10,0} = \exp\left\{-\frac{3}{4}[p_2k(p_2k-1)]^{\frac{1}{3}} p_1 (2b_1)^{\frac{2}{3}}\right\},$$

then under M_0*,*

$$\log B_{10} - \log R_{10,0} = O_p(1).$$

(c) Under M_3*,*

$$\log\left(\frac{B_{10}}{R_{10,1}}\right) \le o(1), \, a.s..$$

With the additional assumption in (b), under M_2

$$\log B_{10} - \log R_{10,0} \le O_p(1).$$

Proof. For Parts (a) and (b), the proof is the same as in the theorem for 1-way ANOVA in Sun et al. (2012). For Part (c), note that in (4.6), the distribution of SST under M_3 is greater than that under M_1, and its distribution under M_2 is greater than that under M_0. $\square$

For B_{20}, the result in Proposition 1 still holds, and since p_2 is fixed,

$$\log\left(\frac{B_{20}}{R_{20}}\right) \xrightarrow{P} 0,$$

if $nb_2 \to \infty$ as $N \to \infty$.

Next, we obtain the approximation for B_{30} under this scenario.

Proposition 3. *Assume that $nb_2 \to \infty$ and $b_1 \to 0$ as $N \to \infty$, and define*

$$Q(t) = \Big(b_2 + \frac{SSR_2}{2(SST - \frac{p_2kt}{1+p_2kt}SSR_1 - SSR_2)} + \frac{p_2-1}{2n}\Big)^{-\frac{p_2}{2}}.$$

(a) Let

$$\begin{aligned} R_{30,3} &= n^{\frac{1-p_2}{2}}\sqrt{\frac{b_2}{\pi}}\Gamma(\frac{p_2}{2})\Big(\frac{SST}{SST-SSR_2}\Big)^{\frac{n-1}{2}}Q(\tilde{g}_1) \\ &\quad \exp\Big\{\frac{p_1}{2}\log\Big[\frac{(1+\tilde{g}_1)^{p_2k}}{1+p_2k\tilde{g}_1}\Big] - \log p_1\Big\}\Big[\frac{4b_1(1+p_2k\tilde{g}_1)^3}{p_2k(p_2k-1)\tilde{g}_1^3(1+\tilde{g}_1)}\Big]^{\frac{1}{2}}, \end{aligned}$$

then when $\tilde{g}_1 > 0$ (under M_1 or M_3),

$$\log\left(\frac{B_{30}}{R_{30,3}}\right) \xrightarrow{P} 0.$$

(b) Assume that $p_1^3b_1^2 \to \infty$ as $N \to \infty$. Let

$$\begin{aligned} R_{30,0} &= n^{\frac{1-p_2}{2}}\sqrt{\frac{b_2}{\pi}}\Gamma(\frac{p_2}{2})\Big(\frac{SST}{SST-SSR_2}\Big)^{\frac{n-1}{2}}Q(0) \\ &\quad \exp\Big\{-\frac{3}{4}[p_2k(p_2k-1)]^{\frac{1}{3}}p_1(2b_1)^{\frac{2}{3}}\Big\}, \end{aligned}$$

then when $\tilde{g}_1 = 0$ (under M_0 or M_2),

$$\log B_{30} - \log R_{30,0} = O_p(1).$$

Proof. We follow the notation in Proposition 1, and

$$B_{30} = n\sqrt{\frac{b_1 b_2}{\pi^2}} \int_0^\infty \int_0^\infty w_1^{\frac{p_1}{2}-1} w_2^{\frac{p_2}{2}-1} \exp\{-nh(w_1, w_2)\} T(w_1, w_2) dw_1 dw_2,$$

We proved that as $N \to \infty$,

$$\frac{\int_\epsilon^\infty w_2^{\frac{p_2}{2}-1} \frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}} dw_2}{\int_0^\epsilon w_2^{\frac{p_2}{2}-1} \frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}} dw_2} \to 0 \tag{4.41}$$

uniformly w.r.t. w_1. Denote the denominator as $J_1(w_1)$, and define

$$H_1(w_1) = T(w_1, 0) \exp[-nh(w_1, 0)] \frac{\Gamma(\frac{p_2}{2})}{[nh_2'(w_1, 0)]^{p_2/2}}.$$

We proved that

$$\frac{J_1(w_1)}{H_1(w_1)} \geq 1 - o(1),$$

where the speed that the right hand side tends to 1 does not depend on w_1. On the other hand,

$$\begin{aligned} \frac{J_1(w_1)}{H_1(w_1)} &\leq \frac{T(w_1, \epsilon)}{T(w_1, 0)} \left[\frac{h_2'(w_1, 0)}{h_2'(w_1, \epsilon)}\right]^{p_2/2} \\ &\leq (1+\epsilon)^{p_2} \left(1 + \frac{\epsilon}{1+\epsilon} \frac{SSR_2}{SSE^*}\right)^{\frac{p_2+1}{2}}. \end{aligned}$$

Therefore, as $N \to \infty$,

$$\frac{\int_0^\infty w_2^{\frac{p_2}{2}-1} \frac{T(w_1,w_2)}{\exp\{nh(w_1,w_2)\}} dw_2}{H_1(w_1)} \xrightarrow{P} 1$$

uniformly w.r.t. w_1.

Therefore, B_{30} is asymptotically equivalent (in probability) to

$$
\begin{aligned}
& n\sqrt{\frac{b_1 b_2}{\pi^2}} \int_0^\infty w_1^{\frac{p_1}{2}-1} T(w_1,0)\exp[-nh(w_1,0)] \frac{\Gamma(\frac{p_2}{2})}{[nh_2'(w_1,0)]^{p_2/2}} dw_1 \\
= \; & n^{1-\frac{p_2}{2}} \sqrt{\frac{b_1 b_2}{\pi^2}} \Gamma(\frac{p_2}{2}) \Big(\frac{SST - SSR_2}{SST}\Big)^{-\frac{n-1}{2}} \int_0^\infty \Big(1 - \frac{1}{1+w_1}\frac{SSR_1}{SST - SSR_2}\Big)^{-\frac{n-1}{2}} \\
& \Big(b_2 + \frac{SSR_2}{2(SST - \frac{SSR_1}{1+w_1} - SSR_2)} + \frac{p_2 - 1}{2n}\Big)^{-\frac{p_2}{2}} w_1^{\frac{p_1}{2}-1} (1+w_1)^{-\frac{p_1-1}{2}} e^{-b_1 n w_1} dw_1 \\
= \; & n^{\frac{1-p_2}{2}} \sqrt{\frac{b_2}{\pi}} \Gamma(\frac{p_2}{2}) \Big(\frac{SST - SSR_2}{SST}\Big)^{-\frac{n-1}{2}} \int_0^\infty \Big(1 - \frac{g_1}{1+g_1}\frac{SSR_1}{SST - SSR_2}\Big)^{-\frac{n-1}{2}} \\
& \Big(b_2 + \frac{SSR_2}{2(SST - \frac{g_1}{1+g_1} SSR_1 - SSR_2)} + \frac{p_2 - 1}{2n}\Big)^{-\frac{p_2}{2}} \frac{\mathrm{IG}(g_1; \frac{1}{2}, nb_1)}{(1+g_1)^{\frac{p_1-1}{2}}} dg_1 \\
= \; & n^{\frac{1-p_2}{2}} \sqrt{\frac{b_2}{\pi}} \Gamma(\frac{p_2}{2}) \Big(\frac{SST - SSR_2}{SST}\Big)^{-\frac{n-1}{2}} \int_0^\infty \Big(1 - \frac{p_2 k t}{1+p_2 k t}\frac{SSR_1}{SST - SSR_2}\Big)^{-\frac{n-1}{2}} \\
& \Big(b_2 + \frac{SSR_2}{2(SST - \frac{p_2 k t}{1+p_2 k t} SSR_1 - SSR_2)} + \frac{p_2 - 1}{2n}\Big)^{-\frac{p_2}{2}} \frac{\mathrm{IG}(t; \frac{1}{2}, p_1 b_1)}{(1+p_2 k t)^{\frac{p_1-1}{2}}} dt
\end{aligned}
$$

The last step is obtained by transformation $t = \frac{g_1}{p_2 k}$. The integration part is equal to

$$
\sqrt{\frac{p_1 b_1}{\pi}} \Big(1 + \frac{SSR_1}{SSE^*}\Big)^{\frac{n_1}{2}} \int_0^\infty \exp\{-p_1 G(t)\} S(t) dt,
$$

where

$$
\begin{aligned}
G(t) &= \frac{p_2 k}{2} \log\Big(1 + p_2 k t + \frac{SSR_1}{SSE^*}\Big) - \frac{p_2 k - 1}{2} \log(1 + p_2 k t) + \frac{b_1}{t}, && (4.42) \\
S(t) &= t^{-\frac{3}{2}} \Big(1 + p_2 k t + \frac{SSR_1}{SSE^*}\Big)^{\frac{1}{2}} Q(t). && (4.43)
\end{aligned}
$$

This differs from the theorem for 1-way ANOVA only in $S(t)$, namely the additional term $Q(t)$. Similar to the discussion for 1-way ANOVA, we can partition the integration into $(J_1 + J_0 + J_2 + J_3)$ according to the range of integration. Then the Laplace approximation on J_0 still holds with the additional term $Q(\tilde{g}_1)$. To show that

$J_i/J_0 \to 0$ $(i = 1, 2, 3)$ still holds, just notice that

$$\frac{Q(t)}{Q(\tilde{g}_1)} \leq \frac{Q(0)}{Q(\infty)} = \left(\frac{b_2 + \frac{SSR_2}{2SSE^*} + \frac{p_2-1}{2n}}{b_2 + \frac{SSR_2}{2(SST-SSR_2)} + \frac{p_2-1}{2n}} \right)^{\frac{p_2}{2}} \leq \left(1 + \frac{SSR_1}{SSE^*}\right)^{\frac{p_2}{2}},$$

where the right hand side goes to a constant as $N \to \infty$. □

4.4.2 Consistency of the Bayes factors

Next, we prove the consistency of these Bayes factors one by one. First, as shown in Theorem 6, B_{20} is consistent under both M_0 and M_2.

For B_{10}, similar to the discussion for 1-way ANOVA in Sun et al. (2012), we have the following results.

Theorem 11. *Assume that $b_1 \to 0$ as $N \to \infty$.*

(a) Under M_1,

$$\frac{2}{p_1} \log B_{10} \xrightarrow{a.s.} \log \left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k \tilde{g}_1)} \right], \text{ as } N \to \infty.$$

(b) Under M_0, if $p_1^3 b_1^2 \to \infty$ as $N \to \infty$, then

$$\frac{1}{p_1 b_1^{\frac{2}{3}}} \log B_{10} \xrightarrow{P} -\frac{3}{2^{\frac{4}{3}}} [p_2 k (p_2 k - 1)]^{\frac{1}{3}}, \text{ as } N \to \infty.$$

(c) B_{10} is consistent under both M_0 and M_1.

The consistency of B_{30} is given in the next theorem.

Theorem 12. *Assume that $nb_2 \to \infty$, $b_2 = O(1)$, and $b_1 \to 0$ as $N \to \infty$.*

(a) Further assume that $p_1^3 b_1^2 \to \infty$ as $N \to \infty$. Then under M_0, as $N \to \infty$,

$$\frac{1}{\frac{p_2-1}{2}\log(nb_2) + \frac{3}{2^{4/3}} p_1 b_1^{\frac{2}{3}} [p_2 k (p_2 k - 1)]^{\frac{1}{3}}} \log B_{30} \xrightarrow{P} -1. \tag{4.44}$$

(b) Under M_3, as $N \to \infty$,

$$\frac{2}{n}\log B_{30} \to \frac{1}{p_2 k}\log\left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k\tilde{g}_1)}\right] + \log\left(1+\frac{1}{1+\tilde{g}_1}D\right), \tag{4.45}$$

where

$$D = \begin{cases} \frac{p_2-1}{p_2}\tilde{g}_2, & \text{if } \boldsymbol{\beta}_2 \text{ are fixed effects,} \\ \frac{\tilde{g}_2}{p_2}\chi^2_{p_2-1}, & \text{if } \boldsymbol{\beta}_2 \text{ are random effects,} \end{cases}$$

and the convergence is in probability if $\boldsymbol{\beta}_2$ are fixed effects and in distribution if $\boldsymbol{\beta}_2$ are random effects.

(c) B_{30} is consistent under both M_0 and M_3.

Proof. For Part (a), we just need to show (4.44) for $R_{30,0}$, for which just note that the leading term in $\log R_{30,0}$ is

$$-\frac{p_2-1}{2}\log(nb_2) - \frac{3}{2^{4/3}}p_1 b_1^{\frac{2}{3}}[p_2 k(p_2 k-1)]^{\frac{1}{3}}.$$

For Part (b), note that the leading term in $\log R_{30,3}$ is

$$\frac{p_1}{2}\log\left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k\tilde{g}_1)}\right] + \frac{n-1}{2}\log\left(1+\frac{SSR_2}{SSE^*+SSR_1}\right).$$

In the expression above,

$$\frac{SSR_2}{SSE^*+SSR_1} = \frac{SSR_2}{SSE^*}\frac{1}{1+\frac{SSR_1}{SSE^*}},$$

and according to Lemma 5, as $N \to \infty$,

$$\frac{SSR_1}{SSE^*} \xrightarrow{P} \frac{1}{1-c_1}(\tilde{g}_1 + c_1), \quad \frac{SSR_2}{SSE^*}\begin{cases} \xrightarrow{P} \frac{1}{1-c_1}\frac{p_2-1}{p_2}\tilde{g}_2, & \text{if } \boldsymbol{\beta}_2 \text{ are fixed effects,} \\ \xrightarrow{L} \frac{1}{1-c_1}\frac{\tilde{g}_2}{p_2}\chi^2_{p_2-1}, & \text{if } \boldsymbol{\beta}_2 \text{ are random effects.} \end{cases}$$

Thus, Part (b) is proved, and Part (c) follows from Parts (a) and (b). □

Theorem 13. *Assume that $nb_2 \to \infty$, $b_2 = O(1)$, and $b_1 \to 0$ as $N \to \infty$.*

(a) Under M_1,

$$\frac{2}{(p_2-1)\log(nb_2)} \log B_{31} \xrightarrow{P} -1,$$

so B_{31} is consistent.

(b) Under M_3, B_{31} is consistent.

Proof. For Part (a), under M_1,

$$\begin{aligned} \log B_{31} &= \log R_{30,3} - \log R_{10,1} + o_P(1) \\ &= \log R_{20} + \log\left(\frac{Q(\tilde{g}_1)}{Q(0)}\right) + o_p(1), \\ &= \log R_{20} + O_p(1), \end{aligned}$$

where the leading term is $-\frac{p_2-1}{2}\log(nb_2)$.

For Part (b), note that under M_3,

$$\frac{2}{n}\log B_{31} \geq \frac{2}{n}\log R_{20} + O_p(\frac{1}{n}),$$

where the leading term on the right hand side is

$$\log\left(1+\frac{SSR_2}{SST-SSR_2}\right)\begin{cases} \xrightarrow{P} \log\left(1+\frac{1}{1+\tilde{g}_1}\frac{p_2-1}{p_2}\tilde{g}_2\right), & \text{if } \boldsymbol{\beta}_2 \text{ are fixed effects,} \\ \xrightarrow{L} \log\left(1+\frac{1}{1+\tilde{g}_1}\frac{\tilde{g}_2}{p_2}\chi^2_{p_2-1}\right), & \text{if } \boldsymbol{\beta}_2 \text{ are random effects.} \end{cases}$$

Thus, B_{31} is consistent under M_3. □

Theorem 14. *Assume that $nb_2 \to \infty$, $b_2 = O(1)$, and $b_1 \to 0$ as $N \to \infty$.*

(a) Further assume that $p_1^3 b_1^2 \to \infty$ as $N \to \infty$. Under M_2,

$$\frac{1}{p_1 b_1^{\frac{2}{3}}} \log B_{32} \xrightarrow{P} -\frac{3}{2^{\frac{4}{3}}}[p_2 k(p_2 k - 1)]^{\frac{1}{3}}, \text{ as } N \to \infty.$$

(b) Under M_3,

$$\frac{2}{p_1} \log B_{32} \xrightarrow{P} \log\left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k \tilde{g}_1)}\right], \text{ as } N \to \infty.$$

(c) B_{32} is consistent under both M_2 and M_3.

Proof. Under M_2,

$$\begin{aligned} \log B_{32} &= \log R_{30,0} + O_p(1) - \log R_{20} + o_p(1) \\ &= -\frac{3}{4}[p_2 k(p_2 k - 1)]^{\frac{1}{3}} p_1 (2b_1)^{\frac{2}{3}} + O_p(1), \end{aligned}$$

so Part (a) is proved. Under M_3,

$$\log B_{32} = \log R_{30,3} - \log R_{20} + o_p(1),$$

the leading term of which is $\frac{p_1}{2} \log\left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k \tilde{g}_1)}\right]$, so Part (b) is proved. Part (c) follows from (a) and (b). □

Theorem 15. *Assume that $nb_2 \to \infty$, $b_2 = O(1)$, $b_1 \to 0$ as $N \to \infty$.*

(a) Further assume that $p_1^3 b_1^2 \to \infty$ as $N \to \infty$. Under M_2, B_{21} is consistent.

(b) Under M_1,

$$\frac{2}{p_1} \log B_{21} \xrightarrow{P} \log\left[\frac{(1+\tilde{g}_1)^{p_2 k}}{(1+p_2 k \tilde{g}_1)}\right],$$

and B_{21} is consistent.

Proof. For Part (a), under M_2,

$$\log B_{21} \geq \log R_{20} + o_p(1) - \log R_{10,0} + O_p(1),$$

The leading term on the right hand side is

$$\frac{n-1}{2} \log\Big(1 + \frac{SSR_2}{SST - SSR_2}\Big),$$

since $\frac{3}{4}[p_2k(p_2k-1)]^{\frac{1}{3}} p_1(2b_1)^{\frac{2}{3}} = o(n)$. B_{21} is consistent since under M_2

$$\frac{SSR_2}{SST - SSR_2} \begin{cases} \xrightarrow{P} \frac{p_2-1}{p_2}\tilde{g}_2, & \text{if } \boldsymbol{\beta}_2 \text{ are fixed effects,} \\ \xrightarrow{L} \frac{\tilde{g}_2}{p_2}\chi^2_{p_2-1}, & \text{if } \boldsymbol{\beta}_2 \text{ are random effects.} \end{cases}$$

For Part (b), under M_1,

$$\log B_{21} = \log R_{20} + o_p(1) - \log R_{10,1} + O_p(1),$$

note that the leading term is

$$-\frac{p_2-1}{2}\log(nb_2) - \frac{p_1}{2}\log\Big[\frac{(1+\tilde{g}_1)^{p_2k}}{(1+p_2k\tilde{g}_1)}\Big],$$

and the first term here is actually of smaller order. □

4.5 Discussion

The results from the previous sections suggest how b_i's in the prior should be chosen in order to obtain consistent Bayes factors for 2-way ANOVA models with main effects only. Specifically, for Scenario I, we could let $p_i = o(nb_i)$ and $b_i = O(1)$ as $N \to \infty$ for $i = 1, 2$. For Scenario II, if only p_1 goes to infinity, we could let $b_1 \to 0$,

$p_1^3 b_1^2 \to \infty$, $p_2 = o(nb_2)$, and $b_2 = O(1)$ as $N \to \infty$. If only p_2 goes to infinity, we could switch the conditions on b_1 and b_2. Note that $p_i = o(nb_i)$ $(i = 1, 2)$ under Scenario I is a sufficient, but not necessary, condition that makes the gamma-type Laplace approximation possible. The Bayes factors might still be consistent when this condition is not true, yet alternative methods are needed for approximating them.

The method 'IZS' in the previous chapters corresponds to $b_1 = b_2 = 1/2$, so it can be applied under Scenario I. Whereas under Scenario II, Sun et al. (2012) proved that 'IZS' leads to inconsistent Bayes factors. The method 'TESS' corresponds to $b_i = 1/(2p_i)$ for $i = 1, 2$, so it satisfies the conditions for Scenario II. For Scenario I, 'TESS' will meet the conditions only when $p_i^2 = o(n)$, which is not always satisfied. However, since the condition for this scenario is not necessary, 'TESS' is likely to still yield consistent Bayes factors (as suggested in the simulation studies in the previous chapter), which needs further discussion. In conclusion, we know that we could apply 'IZS' for Scenario I and 'TESS' for Scenario II. This is not completely satisfactory from a practical point of view since which scenario a real dataset comes from is usually unknown, so a unified treatment for both scenarios is more desirable. We believe that 'TESS' could potential be the choice but need more justification.

Moreover, our current approximation and discussion for 2-way ANOVA models can be easily generalized to m-way ANOVA models with main effects only. For which we can still discuss under 2 scenarios, where the first scenario is still when $p_i = o(n)$ for $i = 1, \ldots, m$, and the second scenario is when $p_i = o(n)$ for all but one i in $\{1, \ldots, m\}$. For the first scenario, the approximating method and results in Proposition 1 can be extended. The techniques in Proposition 3 can be generalized to incorporate the second scenario.

Chapter 5

Future Work

5.1 More on 2-Way ANOVA Models with Main Effects

In Chapter 4, it is proved that for 2-way ANOVA models with main effects, under Scenario I, 'IZS' yields consistent Bayes factors, whereas under Scenario II, the Bayes factors with 'TESS' are consistent. However, this only partially solved the problem. Specifically, in practice, it would be unable to identify the scenario for a set of data, and it is more desirable to have a unified solution for both scenarios. As suggested by the simulation results in Chapter 3, 'TESS' should also work under Scenario I, so 'TESS' is potentially the treatment for both scenarios. To that end, further discussion is needed on the consistency of the Bayes factors with 'TESS' under Scenario I and alternative approximations are needed.

5.2 An Alternative Parametrization for ANOVA Models

When the interaction effects are added into consideration, the Bayes factors for the ANOVA models become rather difficult to deal with. For example, consider M_4 in Example 1 again. The Bayes factor between it and M_0 is

$$B_{40} = \int_0^\infty \int_0^\infty \int_0^\infty (1+g_1+g_3)^{-\frac{p_1-1}{2}} (1+g_2+g_3)^{-\frac{p_2-1}{2}} (1+g_3)^{-\frac{p_1p_2-p_1-p_2+1}{2}}$$
$$\Big(1 - \frac{g_3}{1+g_3}\frac{SSAB}{SST} - \frac{g_1+g_3}{1+g_1+g_3}\frac{SSA}{SST} - \frac{g_2+g_3}{1+g_2+g_3}\frac{SSB}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{3} \pi(g_i) dg_i.$$

The hyper-parameters g_j's are mixed together in the integrand, which complicates the approximation.

A possible solution is to consider an alternative parametrization where the design matrices are orthogonal to the common part of all the models. Consider the general model in (3.3) again.

$$\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_m\boldsymbol{\beta}_m + \boldsymbol{\epsilon}, \tag{5.1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector, for $j = 0, 1, \ldots, m$, $\boldsymbol{X}_j$ is an $n \times p_j$ known design matrix of full column rank, $\boldsymbol{\beta}_j$ is a $p_j \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\epsilon} \sim N_n\big(\mathbf{0}, \sigma^2\boldsymbol{I}_n\big)$.

Let $C(\boldsymbol{X}_j)$ be the column space of $\boldsymbol{X}_j$ (for $j = 0, 1, \ldots, m$). We assume that for $i \neq j \in \{1, \ldots, m\}$,

$$C(\boldsymbol{X}_0) \subsetneq C(\boldsymbol{X}_i), \quad C(\boldsymbol{X}_0) = C(\boldsymbol{X}_i) \cap C(\boldsymbol{X}_j). \tag{5.2}$$

For an ANOVA model with main effects only, where $\boldsymbol{X}_0 = \mathbf{1}_n$, (5.2) is usually satisfied. In this case, the model is actually over-parameterized. We can orthogonalize $\boldsymbol{X}_j$ (for

$j = 1, \ldots, m$) to $\boldsymbol{X}_0$ by considering

$$\boldsymbol{X}_j^* = (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}_j.$$

The obtained $\boldsymbol{X}_j^*$ is not of full column rank (its rank is $(p_j - p_0)$), so we can use the first $(p_j - p_0)$ columns of it to construct an $n \times (p_j - p_0)$ matrix, which we denote as $\widetilde{\boldsymbol{X}}_j$. The model (5.1) under this new parametrization is

$$\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{X}}_1\tilde{\boldsymbol{\beta}}_1 + \cdots + \widetilde{\boldsymbol{X}}_m\tilde{\boldsymbol{\beta}}_m + \boldsymbol{\epsilon}, \tag{5.3}$$

where $\tilde{\boldsymbol{\beta}}_j$'s are the new parameters. Note that there are other ways of defining the design matrices such that they are orthogonal to $\boldsymbol{X}_0$, and the meanings of the corresponding $\boldsymbol{\beta}$ parameters are different.

For this parametrization, the design matrices corresponding to the interaction effects can be defined based on $\widetilde{\boldsymbol{X}}_j$'s. For example, for $j = 1, \ldots, m$, let

$$\widetilde{\boldsymbol{X}}_j = \begin{pmatrix} \boldsymbol{x}'_{j,1} \\ \vdots \\ \boldsymbol{x}'_{j,n} \end{pmatrix},$$

then the design matrix corresponding to the interaction between $\widetilde{\boldsymbol{X}}_{j_1}$ and $\widetilde{\boldsymbol{X}}_{j_2}$ (without loss of generality, suppose $j_1 < j_2$) is

$$\widetilde{\boldsymbol{X}}_{j_1 j_2} = \begin{pmatrix} \boldsymbol{x}'_{j_1,1} \otimes \boldsymbol{x}'_{j_2,1} \\ \vdots \\ \boldsymbol{x}'_{j_1,n} \otimes \boldsymbol{x}'_{j_2,n} \end{pmatrix}.$$

Using similar approach, we can also define the matrices for higher order interactions.

Under this orthogonal parametrization, the marginal likelihood and the Bayes factors become simpler. Consider the balanced q-way full factorial ANOVA models for example. As before, suppose that the j-th factor has p_j levels for $j = 1, \ldots, q$, then with k replicates at each combination of levels, the sample size is $n = k \prod_{j=1}^{q} p_j$. Note that $\boldsymbol{X}_0 = \mathbf{1}_n$, $p_0 = 1$, and $\boldsymbol{P}_0 = \frac{1}{n}\boldsymbol{J}_n$. Under the new parametrization, the design matrix for the main effect of the j-th factor is now

$$\widetilde{\boldsymbol{X}}_j = \otimes_{i=1}^{q} \boldsymbol{W}_i^{(j)} \otimes \mathbf{1}_k, \tag{5.4}$$

where $\boldsymbol{W}_i^{(j)} = \mathbf{1}_{p_i}$ when $i \neq j$, and $\boldsymbol{W}_i^{(j)}$ is the first $(p_i - 1)$ columns of $(\boldsymbol{I}_{p_i} - \frac{1}{p_i} J_{p_i})$ (denoted as $\boldsymbol{O}_i$) when $i = j$.

Next, we show that the design matrix corresponding to the interaction effects of the j_1-th and j_2-th factors is

$$\widetilde{\boldsymbol{X}}_{j_1 j_2} = \otimes_{i=1}^{q} \boldsymbol{W}_i^{\{j_1, j_2\}} \otimes \mathbf{1}_k, \tag{5.5}$$

where $\boldsymbol{W}_i^{\{j_1,j_2\}} = \mathbf{1}_{p_i}$ if $i \neq j_1, j_2$, and $\boldsymbol{W}_i^{\{j_1,j_2\}} = \boldsymbol{O}_i$ if $i = j_1$ or j_2. We only prove (5.5) for the first row, and the rest is similar. On the one hand, the first row of $\widetilde{\boldsymbol{X}}_j$ is

$$\begin{aligned}
& (1, 0, \ldots, 0)_{1\times n} \widetilde{\boldsymbol{X}}_j \\
= \;& \Big(\otimes_{i=1}^{q} (1, 0, \ldots, 0)_{1\times p_i} \otimes (1, 0, \ldots, 0)_{1\times k} \Big) \Big(\otimes_{i=1}^{q} \boldsymbol{W}_i^{(j)} \otimes \mathbf{1}_k \Big) \\
= \;& \otimes_{i=1}^{q} \Big((1, 0, \ldots, 0)_{1\times p_i} \boldsymbol{W}_i^{(j)} \Big) \otimes \Big((1, 0, \ldots, 0)_{1\times k} \mathbf{1}_k \Big) \\
= \;& (1, 0, \ldots, 0)_{1\times p_j} \boldsymbol{O}_j.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& (1, 0, \ldots, 0)_{1\times n} \Big(\otimes_{i=1}^{q} \boldsymbol{W}_i^{\{j_1,j_2\}} \otimes \mathbf{1}_k \Big) \\
= \;& \otimes_{i=1}^{q} \Big((1, 0, \ldots, 0)_{1\times p_i} \boldsymbol{W}_i^{\{j_1,j_2\}} \Big) \otimes \Big((1, 0, \ldots, 0)_{1\times k} \mathbf{1}_k \Big) \\
= \;& \Big((1, 0, \ldots, 0)_{1\times p_{j_1}} \boldsymbol{O}_{j_1} \Big) \otimes \Big((1, 0, \ldots, 0)_{1\times p_{j_2}} \boldsymbol{O}_{j_2} \Big),
\end{aligned}$$

which is equal to the first row of $\widetilde{\boldsymbol{X}}_{j_1 j_2}$. Therefore, (5.5) is proved. Moreover, for $\tau \subseteq \{1, \ldots, q\}$, the design matrix for the interaction effects corresponding to τ is

$$\widetilde{\boldsymbol{X}}_\tau = \otimes_{i=1}^q \boldsymbol{W}_i^{(\tau)} \otimes \boldsymbol{1}_k, \tag{5.6}$$

where $\boldsymbol{W}_i^{(\tau)} = \boldsymbol{1}_{p_i}$ when $i \notin \tau$ and $\boldsymbol{W}_i^{(\tau)} = \boldsymbol{O}_i$ when $i \in \tau$.

Recall that we define $\boldsymbol{\Gamma}^* = \{\{1\}, \ldots, \{q\}, \{1, 2\}, \ldots, \{1, \ldots, q\}\}$, the model with all the interaction effects is

$$\boldsymbol{y} = \boldsymbol{1}_n \mu + \sum_{\tau \in \boldsymbol{\Gamma}^*} \widetilde{\boldsymbol{X}}_\tau \tilde{\boldsymbol{\beta}}_\tau + \boldsymbol{\epsilon}, \tag{5.7}$$

and we denote this model as $M_{\boldsymbol{\Gamma}^*}$. By indexing the model space with the subsets of $\boldsymbol{\Gamma}^*$, we can also define the sub-models of $M_{\boldsymbol{\Gamma}^*}$. For $\boldsymbol{\Lambda} \subseteq \boldsymbol{\Gamma}^*$, let $M_{\boldsymbol{\Lambda}}$ be the model with all the main and interaction effects corresponding to the terms in $\boldsymbol{\Lambda}$:

$$\boldsymbol{y} = \boldsymbol{1}_n \mu + \sum_{\tau \in \boldsymbol{\Lambda}} \widetilde{\boldsymbol{X}}_\tau \tilde{\boldsymbol{\beta}}_\tau + \boldsymbol{\epsilon}. \tag{5.8}$$

Theorem 16. *Given the hyper-parameters $\boldsymbol{g}$, the Bayes factor between model (5.8) and the null model $\boldsymbol{M}_\emptyset$ is*

$$B_{\boldsymbol{\Lambda}:\emptyset}(\boldsymbol{y} \mid \boldsymbol{g}) = \prod_{\tau \in \boldsymbol{\Lambda}} (1 + g_\tau)^{-\frac{p_\tau}{2}} \Big[1 - \sum_{\tau \in \boldsymbol{\Lambda}} \frac{g_\tau}{g_\tau + 1} \frac{\boldsymbol{y}' \widetilde{\boldsymbol{P}}_\tau \boldsymbol{y}}{SST}\Big]^{-\frac{n-1}{2}}, \tag{5.9}$$

where for $\tau \subseteq \{1, \ldots, q\}$, $\widetilde{\boldsymbol{P}}_\tau$ is the projection matrix for $\widetilde{\boldsymbol{X}}_\tau$, and $p_\tau = \prod_{j \in \tau}(p_j - 1)$.

Proof. Clearly, for $\tau \subseteq \{1, \ldots, q\}$,

$$\begin{aligned} \widetilde{\boldsymbol{P}}_\tau &= \widetilde{\boldsymbol{X}}_\tau (\widetilde{\boldsymbol{X}}_\tau' \widetilde{\boldsymbol{X}}_\tau)^{-1} \widetilde{\boldsymbol{X}}_\tau' \\ &= \otimes_{j=1}^q \Big[\boldsymbol{W}_j^{(\tau)} (\boldsymbol{W}_j^{(\tau)\prime} \boldsymbol{W}_j^{(\tau)})^{-1} \boldsymbol{W}_j^{(\tau)\prime}\Big] \otimes \frac{1}{k} \boldsymbol{J}_k. \end{aligned} \tag{5.10}$$

For $j = 1, \ldots, q$, all the columns of $\boldsymbol{O}_j$ are centered at zero,

$$\boldsymbol{O}_j(\boldsymbol{O}_j'\boldsymbol{O}_j)^{-1}\boldsymbol{O}_j'\frac{1}{p_j}\boldsymbol{J}_{p_j} = \frac{1}{p_j}\boldsymbol{J}_{p_j}\boldsymbol{O}_j(\boldsymbol{O}_j'\boldsymbol{O}_j)^{-1}\boldsymbol{O}_j' = \boldsymbol{0}. \tag{5.11}$$

Therefore, $\forall \tau \neq \tau^* \in \boldsymbol{\Gamma}^*$, $\widetilde{\boldsymbol{P}}_\tau \widetilde{\boldsymbol{P}}_{\tau^*} = \widetilde{\boldsymbol{P}}_\tau \widetilde{\boldsymbol{P}}_{\tau^*} = \boldsymbol{0}$, so the commutativity condition is satisfied. Theorem 2 can be applied, where $\forall \boldsymbol{\gamma} \subseteq \boldsymbol{\Gamma}^*$, if $|\boldsymbol{\gamma}| \geq 2$, $\boldsymbol{P}_{\boldsymbol{\gamma}} = \boldsymbol{A}_{\boldsymbol{\gamma}} = \boldsymbol{0}$; if $\boldsymbol{\gamma} = \{\tau\}$, $\boldsymbol{P}_{\boldsymbol{\gamma}} = \boldsymbol{A}_{\boldsymbol{\gamma}} = \widetilde{\boldsymbol{P}}_\tau$, and $p_{\boldsymbol{\gamma}} = p_\tau = \prod_{j \in \tau}(p_j - 1)$. The conclusion follows. □

Specially, it can be shown that

$$\boldsymbol{O}_j(\boldsymbol{O}_j'\boldsymbol{O}_j)^{-1}\boldsymbol{O}_j' = \boldsymbol{I}_{p_j} - \frac{1}{p_j}\boldsymbol{J}_{p_j}, \tag{5.12}$$

then $\boldsymbol{y}'\boldsymbol{P}_\tau \boldsymbol{y}$ is the sum of squares of interaction corresponding to τ. The Bayes factors can be further simplified. For example, for the 2-way ANOVA model with interaction effects discussed earlier in this section, the Bayes factor becomes

$$\begin{aligned} B_{40} = &\int_0^\infty \int_0^\infty \int_0^\infty (1+g_1)^{-\frac{p_1-1}{2}}(1+g_2)^{-\frac{p_2-1}{2}}(1+g_3)^{-\frac{p_1p_2-p_1-p_2+1}{2}} \\ &\Big(1 - \frac{g_3}{1+g_3}\frac{SSAB}{SST} - \frac{g_1}{1+g_1}\frac{SSA}{SST} - \frac{g_2}{1+g_2}\frac{SSB}{SST}\Big)^{-\frac{n-1}{2}} \prod_{i=1}^{3}\pi(g_i)dg_i. \end{aligned} \tag{5.13}$$

The hyper-parameters g_i (i=1,2,3) are now separated in the integrand. Then approximations could be obtained similar to Chapter 4.

Also, note that the choice of $\boldsymbol{O}_j$ is not unique, and the derivation above except (5.12) relies only on the fact that $\boldsymbol{1}_{p_j}'\boldsymbol{O}_j = \boldsymbol{0}$, so we can choose other orthogonal parametrization and still have (5.9). For example, for $j = 1, \ldots, q$, we can also let

$$\boldsymbol{O}_j = \begin{pmatrix} \boldsymbol{I}_{(p_j-1)} \\ -\boldsymbol{1}_{(p_j-1)}' \end{pmatrix}. \tag{5.14}$$

Bibliography

Badar, M. G. & Priest, A. M. (1982), Statistical aspects of fiber and bundle strength in hybrid composites, *in* T. Hayashi, K. Kawata & S. Umekawa, eds, 'Progress in Science and Engineering Composites', ICCM-IV, pp. 1129 – 1136.

Barndorff-Nielsen, O. E. (1983), 'On a formula for the distribution of the maximum likelihood estimator', *Biometrika* **70**, 343–365.

Barndorff-Nielsen, O. E. & Cox, D. R. (1994), *Inference and Asymptotics*, Chapman & Hall, London.

Basu, A. P. (1985), Estimation of the reliability of complex system - a survey., *in* E. J. Dudewicz, ed., 'The Frontiers of Modern Statistical Inference Procedures', American Science Press, Columbus, pp. 271–287.

Bayarri, M. J. & García-Donato, G. (2007), 'Extending conventional priors for testing general hypotheses in linear models', *Biometrika* **94**(1), 135–152.

Berger, J. O., Bayarri, M. J. & Pericchi, L. R. (2010), The effective sample size, Working document, Statistical and Applied Mathematical Sciences Institute.

Berger, J. O. & Bernardo, J. M. (1992), On the development of reference priors (with discussion)., *in* J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, pp. 35–60.

Berger, J. O., Bernardo, J. M. & Sun, D. (2009), 'The formal definition of reference priors.', *Annals of Statistics* **37**(2), 905–938.

Berger, J. O., Ghosh, J. K. & Mukhopadhyay, N. (2003), 'Approximations and consistency of Bayes factors as model dimension grows', *Journal of Statistical Planning and Inference* **112**, 241–258.

Berger, J. O. & Pericchi, L. R. (1996), 'The intrinsic Bayes factor for model selection and prediction', *Journal of the American Statistical Association* **91**(433), 109–122.

Berger, J. O. & Pericchi, L. R. (2001), Objective Bayesian methods for model selection: introduction and comparison, *in* P. Lahiri, ed., 'Model Selection', Vol. 38 of *IMS Lecture Notes- Monograph Series*, Institute of Mathematical Statistics, pp. 135–207.

Bernardo, J. M. (1979), 'Reference posterior distributions for Bayesian inference (with discussion)', *Journal of the Royal Statistical Society: Series B* **41**, 113–147.

Casella, G., Girón, F. J., Martínez, M. L. & Moreno, E. (2009), 'Consistency of Bayesian procedures for variable selection', *The Annals of Statistics* **37**(3), 1207–1228.

Deltell, A. F. (2011), Objective Bayes Criteria for Variable Selection, PhD thesis, Universidad de València.

Dickey, J. (1971), 'The weighted likelihood ratio, linear hypotheses on normal location parameters', *The Annals of Statistics* **42**, 204–223.

Enis, P. & Geisser, S. (1971), 'Estimation of the probability that $Y < X$', *Journal of the American Statistical Association* **66**, 162–168.

Fernández, C., Ley, E. & Steel, M. F. J. (2001), 'Benchmark priors for Bayesian model averaging', *Journal of Econometrics* **100**(2), 381–427.

Ferrari, S. L. P., Ferreira da Silva, M. & Cribari-Neto, F. (2007), 'Adjusted profile likelihoods for the Weibull shape parameter', *Journal of Statistical Computation and Simulation* **77**, 531–548.

Ferreira da Silva, M., Ferrari, S. L. P. & Cribari-Neto, F. (2008), 'Improved likelihood inference for the shape parameter in Weibull regression', *Journal of Statistical Computation and Simulation* **78**, 789–811.

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

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

George, E. I. & McCulloch, R. E. (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **88**(423), 881–889.

Ghosh, M. & Sun, D. (1998), Recent development of Bayesian inference for stress-strength models, *in* 'Frontier of Reliability', World Scientific, pp. 143–158.

Guo, R. & Speckman, P. L. (2009), Bayes factor consistency in linear models, Working document, Department of Statistics, University of Missouri-Columbia.

Hald, A. (1952), *Statistical Theory with Engineering Applications*, Wiley, New York.

Hanagal, D. (2003), 'Estimation of system reliability in multicomponent series stress-strength models', *Journal of Indian Statistical Association* **41**, 1–7.

Hedayat, A., Sloane, N. J. A. & Stufken, J. (1999), *Orthogonal Arrays: Theory and Applications*, Springer Verlag.

Jefferys, W. & Berger, J. (1992), 'Ockham's razor and Bayesian analysis', *American Scientist* **80**, 64–72.

Jeffreys, H. (1961), *Theory of Probability*, Oxford University Press, New York.

Johnson, R. A. (1988), Stress-strength models for reliability, *in* P. Krishnaiah & C. Rao, eds, 'Quality Control and Reliability', Vol. 7 of *Handbook of Statistics*, Elsevier, pp. 27 – 54.

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

Kotz, S., Lumelskii, Y. & Pensky, M. (2003), *The Stress-Strength Model and its Generalizations: Theory and Applications*, World Scientific, Singapore.

Kovacs, A. M. (1996), The Competing Risks Weibull Model for the Strength of Single Fibers, PhD thesis, University of Delaware.

Krishnamoorthy, K. & Lin, Y. (2010), 'Confidence limits for stress-strength reliability involving Weibull models', *Journal of Statistical Planning and Inference* **140**, 1754–1764.

Kundu, D. & Gupta, R. D. (2006), 'Estimation of $P[Y < X]$ for Weibull distribution', *IEEE Transactions on Reliability* **55**(2), 270–280.

Liang, F., Paulo, R., Molina, G., Clyde, M. A. & Berger, J. O. (2008), 'Mixtures of g-priors for Bayesian variable selection', *Journal of the American Statistical Association* **103**(481), 410–423.

Maruyama, Y. & George, E. I. (2011), 'Fully Bayes factors with a generalized g-prior', *Annals of Statistics* **39**(5), 2740–2765.

Moreno, E., Girón, F. J. & Casella, G. (2010), 'Consistency of objective Bayes factors as the model dimension grows', *Annals of Statistics* **38**(4), 1937–1952.

Pace, L. & Salvan, A. (2006), 'Adjustments of the profile likelihood from a new perspective', *Journal of Statistical Planning and Inference* **136**, 3554–3564.

Pérez, J. M. & Berger, J. O. (2002), 'Expected-posterior prior distributions for model selection', *Biometrika* **89**(3), 491.

Raftery, A. E., Madigan, D. & Hoeting, J. A. (1997), 'Bayesian model averaging for linear regression models', *Journal of the American Statistical Association* **92**(437), 179–191.

Reid, N., Mukerjee, R. & Fraser, D. A. S. (2003), 'Some aspects of matching priors', *Lecture Notes-Monograph Series* **42**, pp. 31–43.

Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. (2012), 'Default Bayes factors for ANOVA designs', *Journal of Mathematical Psychology* **56**, 356–374.

Scott, J. G. & Berger, J. O. (2010), 'Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem', *The Annals of Statistics* **38**, 2587–2619.

Severini, T. A. (1998), 'Likelihood functions for inference in the presence of a nuisance parameter', *Biometrika* **85**, 507–522.

Severini, T. A. (2000), *Likelihood Methods in Statistics*, Oxford University Press.

Severini, T. A. (2007), 'Integrated likelihood functions for non-Bayesian inference', *Biometrika* **94**, 529–542.

Sun, D., Ghosh, M. & Basu, A. P. (1998), 'Bayesian analysis for a stress-strength system under noninformative priors', *The Canadian Journal of Statistics* **26**(2), 323–332.

Sun, D., Speckman, P. L., Liu, F. & Rouder, J. N. (2012), 'Bayes factor consistency for one-way ANOVA', *submitted* .

Surles, J. G. & Padgett, W. J. (1998), 'Inference for $P(Y < X)$ in the Burr type X model', *Journal of Applied Statistical Sciences* **7**, 225–238.

Surles, J. G. & Padgett, W. J. (2001), 'Inference for reliability and stress-strength for a scaled Burr type X distribution', *Lifetime Data Analysis* **7**, 187–200.

Thompson, R. D. & Basu, A. P. (1993), Bayesian reliability of stress-strength systems, *in* A. P. Basu, ed., 'Advances in Reliability', Elsevier Science Publishers, pp. 411–421.

Vandaele, W. (1978), Participation in illegitimate activities: Ehrlich revisited, *in* A. Blumstein, J. Cohen & D. Nagin, eds, 'Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates', National Academy of Sciences, pp. 270–335.

Ventura, L., Cabras, S. & Racugno, W. (2009), 'Prior distributions from pseudo-likelihoods in the presence of nuisance parameters', *Journal of the American Statistical Association* **104**(486), 768–774.

Ventura, L. & Racugno, W. (2011), 'Recent advances on Bayesian inference for $P(X < Y)$', *Bayesian Analysis* **6**(3), 411–428.

Verdinelli, I. & Wasserman, L. (1995), 'Computing Bayes factors using a generalization of the Savage-Dickey density ratio', *Journal of the American Statistical Association* **90**(430), 614–618.

Wasserman, L. (2000), 'Asymptotic inference for mixture models using data-dependent priors', *Journal of the Royal Statistical Society: Series B* **62**, 159–180.

Wood, H., Steinour, H. H. & Starke, H. R. (1932), 'Effect of composition of portland cement on heat evolved during hardening', *Industrila and Engineering Chemistry* **24**(11), 1207–1214.

Zellner, A. (1986), On assessing prior distributions and Bayesian regression analysis with g-prior distributions, *in* P. K. Goel & A. Zellner, eds, 'Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti', Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 233–243.

Zellner, A. & Siow, A. (1980), Posterior odds ratios for selected regression hypothesis, *in* J. M. Bernardo, M. H. Degroot, D. V. Lindley & A. F. M. Smith, eds, 'Bayesian Statistics: Proceedings of the First International meeting held in Valencia', University of Valencia Press, pp. 585–603.

VITA

Xiaoyi Min was born on April 26, 1986 in Haimen, a small town in the Jiangsu province of China. After graduating from Peking University with two Bachelor's degrees in Statistics and Economics in 2007, Xiaoyi entered the graduate program in Statistics at the University of Missouri and began research with Dr. Dongchu Sun. He expects to receive a degree of Doctor of Philosophy in December 2012. From January 2013, he will work as a Postdoctoral Associate in the Yale School of Public Health.