Ann Inst Stat Math (2016) 68:877–903
DOI 10.1007/s10463-015-0518-9

Bayesian model selection for a linear model with grouped covariates

Xiaoyi Min[1] · **Dongchu Sun**[2,3]

Received: 9 September 2012 / Revised: 30 January 2015 / Published online: 1 May 2015

Abstract Model selection for normal linear regression models with grouped covariates is considered under a class of Zellner's 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 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.

Keywords ANOVA models · Bayes factor · Consistency · Marginal likelihood · Zellner's g-prior

Electronic supplementary material The online version of this article (doi:10.1007/s10463-015-0518-9) contains supplementary material, which is available to authorized users.

Most of the work was done while Xiaoyi Min was a graduate student at the University of Missouri. This research was supported in part by NSF Grants DMS-1007874 SES-1024080, SES-1260806, and NIH Grant R01DA016750.

✉ Xiaoyi Min
xiaoyi.min@yale.edu

Dongchu Sun
SunD@Missouri.edu

1 Department of Biostatistics, Yale University, 300 George Street Suite 523, New Haven, CT 06511, USA

2 Department of Statistics, University of Missouri, Columbia, MO 65211, USA

3 School of Finance and Statistics, East China Normal University, Shanghai 200241, China

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 and 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 and 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 as $m_i(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}_i} f_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)\pi_i(\boldsymbol{\theta}_i)\mathrm{d}\boldsymbol{\theta}_i$. The Bayes factor of M_i to M_j $(i, j = 1, \ldots, r)$ is defined as $B_{ij} = m_i(\boldsymbol{y})/m_j(\boldsymbol{y})$.

A common and important issue in using the Bayes factor approach is the choice of prior distributions on the parameters. When it comes to linear models, priors need to be selected for both the regression coefficients and the variances. A large amount of literature already addresses this problem. In general, Berger and 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 and Pericchi 2001) introduced by Jeffreys (1961), which was further developed by Zellner and 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 and 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 compared fixed g-priors with mixtures of g-priors. Other priors that have been proposed include intrinsic priors (Berger and Pericchi 1996) and expected posterior priors (Pérez and Berger 2002).

In this paper, we consider a general class of linear models whose covariates could be divided into different groups. This model structure is useful when several covariates and their corresponding coefficients might be of interest jointly rather than separately. Particularly, in ANOVA models, we usually ask whether a certain factor "accounts" for the variations among the responses, and thus coefficients corresponding to different levels of the factor are of interest together. In regression models, different covariates might be related and could be considered together. For example, in genome-wide association studies, different genetic variants might correspond to a same gene on human chromosome and could be regarded as a group. The model we consider can be formulated as follows:

$$\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_m\boldsymbol{\beta}_m + \boldsymbol{\epsilon}, \tag{1}$$

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 $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

We also define sub-models of (1). For $\boldsymbol{\gamma} \subseteq \{1, \ldots, m\}$, let M_γ be $\boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \sum_{j\in\boldsymbol{\gamma}} \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{\epsilon}$. Note that $\boldsymbol{\beta}_0$ represents the common parameters under different models such as the intercept, and $\boldsymbol{\beta}_j$ ($j = 1, \ldots, m$) represent other groups of coefficients potentially useful. To compare different sets of covariates and to decide whether certain groups of covariates are important in the model, we need to find the pairwise Bayes factors of the corresponding models, and the key issue in finding the Bayes factor for any two sub-models, say $B_{\boldsymbol{\gamma\gamma}'}$ for $M_{\boldsymbol{\gamma}}$ and $M_{\boldsymbol{\gamma}'}$ ($\boldsymbol{\gamma}, \boldsymbol{\gamma}' \subseteq \{1, \ldots, m\}$), is assigning appropriate priors on the parameters σ^2 and $\boldsymbol{\beta}_j$ for $j = 0, \ldots, m$.

One of the motivations of this paper is to solve the problems traditional methods such as Zellner's g-prior face when comparing model (1) and its sub-models. In the first place, 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 and the interaction effects of two factors A and B (denoted as M_4), where A has p_1 levels, B has p_2 levels, and each combination of levels has k replicates. Let

$$\boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{X}_3\boldsymbol{\beta}_3 + \boldsymbol{\epsilon}, \tag{2}$$

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, $\boldsymbol{X}_0 = \boldsymbol{1}_n$, $\boldsymbol{X}_1 = \boldsymbol{I}_{p_1} \otimes \boldsymbol{1}_{p_2} \otimes \boldsymbol{1}_k$, $\boldsymbol{X}_2 = \boldsymbol{1}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{1}_k$, and $\boldsymbol{X}_3 = \boldsymbol{I}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \boldsymbol{1}_k$, and $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. We are often interested in comparing M_4 and its sub-models

$$\begin{aligned}
M_0 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{\epsilon},\\
M_1 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon},\\
M_2 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon},\\
M_3 &: \boldsymbol{y} = \boldsymbol{X}_0\mu + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.
\end{aligned}$$

For M_4, the design matrix $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3)$ is not of full column rank, and thus 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.

Second, Zellner's g-prior for a model in (1) 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 $\boldsymbol{\beta}_j$ are limited to changing at the same scale as g changes. In this paper, we try to bring more flexibility to the priors 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 (1), 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 (2) are often considered to be random effects. Moreover, if fixed effects and random effects both exist in a linear model (1), the model is called a linear mixed model. Different definitions for fixed and random effects exist (Gelman 2005), one of which is to consider effects fixed when the effects themselves are of interest and consider them random when the population underlying these effects is of interest. In many real-life applications, however, it is still not easy to decide whether an effect is fixed or random (Searle et al. 1992). 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 and 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. (2010) 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 paper is organized as follows. In Sect. 2, the marginal likelihood of model (1) is derived under the proposed prior, and a simpler closed-form expression is calculated under the commutativity condition. In Sect. 3, the result from Sect. 2 is demonstrated by the application to two special cases. In Sect. 4, the scheme for computing the Bayes factors under the proposed prior is given. In Sect. 5, the performance of the proposed prior is demonstrated using simulation studies. In Sect. 6, the proposed method is applied to two real data studies. Finally, we conclude with stating our findings.

2 Main results

2.1 The proposed priors

Consider the model of (1) and its sub-models. 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 $\boldsymbol{\beta}_j$ are fixed effects, we choose the right Haar prior for the common parameters $(\boldsymbol{\beta}_0, \sigma^2)$,

$$\pi(\boldsymbol{\beta}_0, \sigma^2) = \frac{1}{\sigma^2}. \tag{3}$$

As discussed by Bayarri et al. (2012) and Berger et al. (1998), the right Haar prior, although improper, could and should be used for the common parameters since it satisfies the exact predictive matching criterion as defined in the aforementioned papers. Bayarri et al. (2012) also justified the use of the usually improper right Haar prior by showing that the corresponding Bayes factor is equal to the limit of Bayes factors obtained from any series of proper priors approaching the right Haar prior itself.

As we have discussed earlier, we assign separate g_j to control $\boldsymbol{\beta}_j$ following Zellner's g-prior as follows

$$\boldsymbol{\beta}_j \mid \sigma^2, g_j \overset{\text{indep}}{\sim} N_{p_j}(\mathbf{0}, g_j\sigma^2(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}). \tag{4}$$

Different hyper-priors could be used on g_j in addition to (4) to introduce different joint distributions for the parameters. In particular, if the different groups of covariates are considered independent, we could assign independent inverse-gamma priors on g_j as

$$g_j \overset{\text{indep}}{\sim} \text{Inv-Gamma}(1/2, n/2), \tag{5}$$

which leads to an independent version of the Zellner–Siow prior for the conditional prior $\boldsymbol{\beta}_j$ given $(\boldsymbol{\beta}_0, \sigma^2)$,

$$[\boldsymbol{\beta}_j \mid \boldsymbol{\beta}_0, \sigma^2] = \frac{\Gamma\left(\frac{p_j+1}{2}\right)|X_j'X_j|^{\frac{1}{2}}}{\pi^{\frac{p_j+1}{2}}(n\sigma^2)^{\frac{p_j}{2}}}\left(1 + \frac{1}{n\sigma^2}\boldsymbol{\beta}_j'X_j'X_j\boldsymbol{\beta}_j\right)^{-\frac{p_j+1}{2}}. \tag{6}$$

Other hyper-priors on g_j could be used as the second stage of the hierarchical structure (5) (Liang et al. 2008; Maruyama and George 2011; Guo and Speckman 2009). In particular, Berger et al. (2014) 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 $j = 1, \ldots, m$).

Admittedly, the independence structure in (4) is quite restrictive and should ideally be used only when independence between different groups of $\boldsymbol{\beta}_j$ can be assumed. We note that by specifying dependent hyper-priors for g_j, we can give dependent priors for $\boldsymbol{\beta}_j$ conditional on $(\boldsymbol{\beta}_0, \sigma^2)$ despite that the prior covariances between different groups of $\boldsymbol{\beta}_j$ are still restricted to be zero. Moreover, this independence structure proves convenient for our study as it not only eases the elicitation of the prior correlation structure, but also simplifies the analytic derivation and computation. Similar structures were also used in, for example, George and McCulloch (1993) and Park and Casella (2008).

Note also that for a regression model, we could just use the first stage of this hierarchical structure (4) 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 (4) are often a part of the model, with unknown parameter g_j being the ratio between the corresponding variance component and the error variance σ^2. Different prior choices for the second stage of the hierarchical structure are just priors for g_j.

In the following discussion, we first consider the marginal likelihood given g_j under priors (3) and (4). Then, we consider the prior (5) and the modification with 'TESS' as priors on g_j.

For the convenience of calculation, we use the following notation:

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}^{*'})', \ \boldsymbol{\beta}^* = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')', \quad \boldsymbol{g} \ = \ (g_1, g_2, \ldots, g_m), \tag{7}$$

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

$$\boldsymbol{X} = (\boldsymbol{X}_0, \boldsymbol{X}^*), \ \boldsymbol{X}^* = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m), \tag{8}$$

$$\boldsymbol{M} = \mathrm{diag}(\boldsymbol{0}_{p_0\times p_0}, \boldsymbol{M}_1), \ \boldsymbol{M}_1 = \mathrm{diag}\left(\frac{1}{g_1}\boldsymbol{X}_1'\boldsymbol{X}_1, \frac{1}{g_2}\boldsymbol{X}_2'\boldsymbol{X}_2, \ldots, \frac{1}{g_m}\boldsymbol{X}_m'\boldsymbol{X}_m\right). \tag{9}$$

Notice that Zellner's g-prior on $\boldsymbol{\beta}^*$ given (σ^2, g) is $N\left(\boldsymbol{0}, \sigma^2\left(\frac{1}{g}\boldsymbol{X}^{*'}\boldsymbol{X}^*\right)^{-1}\right)$, and our proposed prior (4) on $\boldsymbol{\beta}^*$ given $(\sigma^2, \boldsymbol{g})$ is $N(\boldsymbol{0}, \sigma^2\boldsymbol{M}_1^{-1})$, 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

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

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

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

Theorem 1 *If the prior on* $\boldsymbol{g}$ *is* $\pi(\boldsymbol{g})$*, the marginal likelihood of* $\boldsymbol{y}$ *will be*

$$m(\boldsymbol{y}) = \frac{\Gamma\left(\frac{n-p_0}{2}\right)}{\pi^{\frac{n-p_0}{2}}}\prod_{j=1}^{m}|\boldsymbol{X}_j'\boldsymbol{X}_j|^{1/2}\int\frac{\prod_{j=1}^{m} g_j^{-p_j/2}}{(\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n-p_0}{2}}|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}|^{1/2}}\pi(\boldsymbol{g})\mathrm{d}\boldsymbol{g}, \tag{10}$$

where

$$\boldsymbol{R} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M})^{-1}\boldsymbol{X}'. \tag{11}$$

Using Theorem 1, we can calculate the marginal likelihood for different models of interest. The pairwise ratios between these marginal likelihood will be the Bayes factor which could be used for model comparison. Next, we discuss a special situation where (10) can be further simplified.

2.2 Commutativity assumption

In (10), we need to compute both

$$|X'X + M| \text{ and } (X'X + M)^{-1}. \tag{12}$$

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

$$P_i P_j = P_j P_i, \quad \forall\, i, j, \tag{13}$$

there is a simple form for (12), so we can write a real closed-form expression of the marginal likelihood without inverting the matrix in (12) numerically.

With this commutativity condition, any product of P_i and P_j reduces to $P_i P_j$ (cf. Baksalary et al. 2002). Consequently, we can derive the explicit forms for the inverse and determinant of any linear combinations of the projection matrices which are essential for calculating (12). Rao and Yanai (1979) pointed out that a necessary and sufficient condition for (13) is that $P_i P_j$ is the orthogonal projection onto the intersection of the spaces spanned by X_i and X_j. This might be a strong condition. In particular, for the ANOVA models that motivated us to consider this problem, commutativity usually requires a balanced design. We refer the readers to Baksalary (1987) for more results on the statistical implications of the commutativity condition.

For $\gamma \subseteq \{0, 1, \ldots, m\}$, define

$$P_\gamma = \prod_{j\in\gamma} P_j, \tag{14}$$

$$A_\gamma = \prod_{j\in\gamma} P_j \prod_{j'\in\{0,1,\ldots,m\}\setminus\gamma} (I_n - P_{j'}), \tag{15}$$

$$p_\gamma = \operatorname{rank}(A_\gamma).$$

Condition (13) guarantees that both (14) and (15) are well defined. For convenience, we define the collection of all nonempty subsets of $\{1, \ldots, m\}$ by $\Gamma = \{\{1\}, \{2\}, \ldots, \{m\}, \{1, 2\}, \ldots, \{1, 2, \ldots, m\}\}$, and use Γ as the index set.

By simplifying (12), we have the following main results.

Theorem 2 *Assume the commutativity condition* (13). *Then* $m(y \mid g)$, *given in* (10), *has the expression,*

$$m(y \mid g) = \frac{\Gamma\left(\frac{n-p_0}{2}\right)}{(\pi\, y' R y)^{\frac{n-p_0}{2}} |X_0' X_0|^{1/2}} \prod_{\gamma\in\Gamma} \frac{1}{(1 + \sum_{j\in\gamma} g_j)^{p_\gamma/2}}, \tag{16}$$

where R *defined in* (11) *has the expression,*

$$R = I_n - P_0 + \sum_{\gamma\in\Gamma} u_\gamma (I_n - P_0) P_\gamma, \tag{17}$$

where

$$u_{\boldsymbol{\gamma}} = (-1)^k \sum_{(j_1, j_2, \ldots, j_k)} \left(\frac{g_{j_1}}{1+g_{j_1}} \frac{g_{j_2}}{1+g_{j_1}+g_{j_2}} \cdots \frac{g_{j_k}}{1+g_{j_1}+\cdots+g_{j_k}} \right), \quad (18)$$

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

Note that (16) gives the explicit form of the marginal likelihood in terms of $\boldsymbol{g}$, which is much easier to work with than (10) in the sense that the marginal likelihood given $\boldsymbol{g}$ can be calculated directly without numerically finding the inverse and determinant of a potentially large matrix. Also, the contribution of different g_j is clearly given, which makes possible the further discussion about the choice of hyper-priors for $\boldsymbol{g}$.

3 Special cases

In this section, we consider a balanced complete factorial q-way ANOVA model. The full model, which has all the main effects and all the interactions, and all its sub-models can be compared. The marginal likelihood for each of these models can be derived using Theorem 2. In particular, we show two special cases here, namely the sub-model with main effects only and the full model. We also show that a complete factorial design is not necessary for the commutativity condition by giving an example in which the theorem can be applied to a fractional design.

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 jth factor has p_j levels ($j = 1, \ldots, q$) and each combination of levels has k replicates, then the model has the same form as (1) with $m = q$ and $n = k \prod_{j=1}^{q} p_j$. The design matrices are $\boldsymbol{X}_0 = \mathbf{1}_n$ and $\boldsymbol{X}_j = \otimes_{i=1}^{m} \boldsymbol{U}_i^{(j)} \otimes \mathbf{1}_k$, 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 (13) is valid because $\boldsymbol{P}_i \boldsymbol{P}_j = \boldsymbol{P}_0 = \frac{1}{n} \boldsymbol{J}_n$, $\forall i \neq j = 0, \ldots, q$. The proposed prior is

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

For the expression (17), for $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, we have:

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

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

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

Similarly, we get

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

Therefore, the marginal likelihood function under this situation is

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

Here, $\boldsymbol{y}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{y}$ is the corrected total sum of squares, and for $j = 1, \ldots, q$, $\boldsymbol{y}'\left(\boldsymbol{P}_j - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{y}$ is the sum of squares corresponding to the jth factor. Subsequently, this marginal likelihood could be compared with the marginal likelihood from other models for model comparison, either condition on $\boldsymbol{g}$ or after the integration with respect to the hyper-priors on $\boldsymbol{g}$.

3.2 Balanced completely factorial q-way ANOVA models with all interactions

For a q-way ANOVA model with all main and interaction effects, we assume that p_j, k, and n are defined similarly as in Sect. 3.1. Since $m = 2^q - 1$ in (1), we use the subsets of $\{1, \ldots, q\}$ as the subscripts of the design matrices and covariate vectors. Specifically, index $\emptyset$ is used instead of "0" to refer to the common part in the model (i.e. the intercept), and any nonempty $\tau \subset \{1, \ldots, q\}$ refers to the main or interaction effect corresponding to the elements in τ. Then, the model (1) can be written as

$$\boldsymbol{y} = \boldsymbol{X}_{\emptyset}\boldsymbol{\beta}_{\emptyset} + \sum_{\emptyset\neq\tau\subset\{1,\ldots,q\}} \boldsymbol{X}_{\tau}\boldsymbol{\beta}_{\tau} + \boldsymbol{\epsilon}, \tag{19}$$

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 $\boldsymbol{X}_{\tau} = \otimes_{i=1}^{q}\boldsymbol{V}_i^{(\tau)} \otimes \boldsymbol{1}_k$. Here, $\boldsymbol{V}_i^{(\tau)} = \boldsymbol{I}_{p_i}$ when $i \in \tau$ and $\boldsymbol{V}_i^{(\tau)} = \boldsymbol{1}_{p_i}$ when $i \notin \tau$. Clearly, $\boldsymbol{X}_{\tau}$ is of full column rank and the commutativity condition (13) is valid because $\boldsymbol{P}_{\tau}\boldsymbol{P}_{\tau^*} = \boldsymbol{P}_{\tau^*}\boldsymbol{P}_{\tau} = \boldsymbol{P}_{\tau\cap\tau^*}$, $\forall\tau, \tau^* \subseteq \{1, \ldots, q\}$. The prior is

$$\pi(\boldsymbol{\beta}_{\emptyset}, \sigma^2) = \frac{1}{\sigma^2},$$
$$\boldsymbol{\beta}_\tau \mid \sigma^2, g_\tau \overset{\text{indep}}{\sim} N_{p_\tau}\left(\mathbf{0}, \sigma^2 \frac{g_\tau \prod_{j\in\tau} p_j}{n} \boldsymbol{I}_{p_\tau}\right).$$

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 (1)], and $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma} \neq \emptyset : \boldsymbol{\gamma} \subseteq \boldsymbol{\Gamma}^*\}$.

Theorem 3 *The marginal likelihood for an m-way ANOVA model* (19) *is*

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

where

$$\boldsymbol{R} = (\boldsymbol{I}_n - \boldsymbol{P}_{\emptyset}) + \sum_{\emptyset\neq\tau\subseteq\{1,2,\ldots,q\}} U_\tau (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{\phi}})\boldsymbol{P}_\tau,$$
$$U_\tau = \begin{cases} \sum_{\tau^*\supseteq\tau} (-1)^{|\tau^*|-|\tau|} \frac{1}{1+\sum_{\tau^{**}\supseteq\tau^*} g_{\tau^{**}}}, & \text{if } \tau \subsetneq \{1, \ldots, q\}, \\ -1 + \frac{1}{1+g_{\{1,\ldots,q\}}}, & \text{if } \tau = \{1, \ldots, q\}. \end{cases}$$

Next, we give an illustration of the conclusion of Theorem 3.

Example 1 (*Continued*). This model is a special case of (19) with $q = 2$. The corresponding marginal likelihood is

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

where

$$\begin{aligned} \text{SST} &= \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{y} = \boldsymbol{y}'\left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{y}, \\ \text{SSA} &= \boldsymbol{y}'(\boldsymbol{P}_1 - \boldsymbol{P}_0)\boldsymbol{y} = \frac{1}{p_2 k}\boldsymbol{y}'\left(\left(\boldsymbol{I}_{p_1} - \frac{1}{p_1}\boldsymbol{J}_{p_1}\right) \otimes \boldsymbol{J}_{p_2} \otimes \boldsymbol{J}_k\right)\boldsymbol{y}, \\ \text{SSB} &= \boldsymbol{y}'(\boldsymbol{P}_2 - \boldsymbol{P}_0)\boldsymbol{y} = \frac{1}{p_1 k}\boldsymbol{y}'\left(\boldsymbol{J}_{p_1} \otimes \left(\boldsymbol{I}_{p_2} - \frac{1}{p_2}\boldsymbol{J}_{p_2}\right) \otimes \boldsymbol{J}_k\right)\boldsymbol{y}, \\ \text{SSAB} &= \boldsymbol{y}'(\boldsymbol{P}_3 - \boldsymbol{P}_2 - \boldsymbol{P}_1 + \boldsymbol{P}_0)\boldsymbol{y} \\ &= \frac{1}{k}\boldsymbol{y}'\left(\left(\boldsymbol{I}_{p_1} - \frac{1}{p_1}\boldsymbol{J}_{p_1}\right) \otimes \left(\boldsymbol{I}_{p_2} - \frac{1}{p_2}\boldsymbol{J}_{p_2}\right) \otimes \boldsymbol{J}_k\right)\boldsymbol{y}. \end{aligned}$$

They are the sums of squares of the corrected total, factor A, factor B, and the AB interaction, respectively.

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 2 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,

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

where

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

In the expression above, $\boldsymbol{E}_1 = \boldsymbol{I}_{2^L} \otimes \mathbf{1}_{2^{K-L}}$, 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) and (6).

4 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 Sect. 2, and we give a procedure to obtain the Bayes factors by applying the Savage–Dickey density ratio (Dickey 1971; Verdinelli and Wasserman 1995).

In this section, we again consider comparing the sub-models of model (1) with $\boldsymbol{X}_0\boldsymbol{\beta}_0$ as the common part for all the models. We give an algorithm for computing the Bayes factor between any pair of these sub-models. First, the Bayes factor between two nested models is considered. The full and the reduced models are assumed to be, respectively,

$$M_F: \boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_R\boldsymbol{\beta}_R + \cdots + \boldsymbol{X}_{R+F}\boldsymbol{\beta}_{R+F} + \boldsymbol{\epsilon},$$
$$M_R: \boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_R\boldsymbol{\beta}_R + \boldsymbol{\epsilon},$$

where the settings on $\boldsymbol{y}$, $\boldsymbol{X}_j$, $\boldsymbol{\beta}_j$, and $\boldsymbol{\epsilon}$ are the same as those of (1). Denote the proposed prior for M_F and M_R as π_F and π_R, respectively. We can see that the marginal of π_F on the extra regression coefficients is centered at M_R where they are all zero. The Bayes factor between M_F and M_R is

$$B_{FR} = m_F/m_R, \tag{20}$$

where m_F and m_R are the marginal likelihood functions corresponding to the two models. Alternatively, the Bayes factor can be defined as $B_{FR} = B_{FN}/B_{RN}$, where B_{FN} is the Bayes factor between M_F and the null model $M_N: \boldsymbol{y} = \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}$ and likewise for B_{RN}. This is equivalent to our definition, and we will show that calculation using (20) is preferable since it only involves one MCMC sampling whereas calculating B_{FN}/B_{RN} requires two independent MCMC steps.

Consider a transformation for M_F and π_F:

$$\boldsymbol{\beta}_j^* = \begin{cases} \boldsymbol{\beta}_j, & \text{for } j = 0, \ldots, R, \\ \frac{1}{\sqrt{g_j \sigma}}\boldsymbol{\beta}_j, & \text{for } j = R+1, \ldots, R+F. \end{cases} \tag{21}$$

Then, the transformed model is

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

with the prior,

$$\pi_F^*(\boldsymbol{\beta}_j^* \mid \sigma^2, g_j) = \begin{cases} \pi_F(\boldsymbol{\beta}_j^* \mid \sigma^2, g_j), & \text{for } j = 1, \ldots, R, \\ N_{p_j}(\boldsymbol{\beta}_j^*; \boldsymbol{0}, (\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1}), & \text{for } j = R+1, \ldots, R+F, \end{cases}$$
$$\pi_F^*(g_j) = \pi_F(g_j), \quad \text{for } j = 1, \ldots, R+F,$$
$$\pi_F^*(\beta_0^*, \sigma^2) = \pi_F(\beta_0^*, \sigma^2).$$

Let m_F^* denote the marginal likelihood under M_F^* and π_F^*. Clearly, $m_F = m_F^*$ and the priors of $\boldsymbol{\beta}_j^*$ do 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. Dickey (1971) showed that if $\pi_F^*(\boldsymbol{\beta}_0^*, \ldots, \boldsymbol{\beta}_R^*, \sigma^2 | \boldsymbol{\beta}_{R+1} = \boldsymbol{0}, \ldots, \boldsymbol{\beta}_{R+F} = \boldsymbol{0}) =$

$\pi_R^*(\boldsymbol{\beta}_0^*, \ldots, \boldsymbol{\beta}_R^*, \sigma^2)$, the Bayes factor can be simplified to the ratio of a prior density and a posterior density (the Savage–Dickey density ratio) as follows

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

The above-mentioned condition and expression are true since under π_F^*, $\boldsymbol{\beta}_j^*$ $(j = R + 1, \ldots, R + F)$ is independent of the other parameters. The denominator in the Savage–Dickey density ratio 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

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

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 (21). Furthermore, the posterior sample from $\pi_F(\cdot \mid \boldsymbol{y})$ can be easily obtained via Gibbs sampling (Gelfand and Smith 1990). In the same way as (7), (8), and (9), we define $\boldsymbol{\beta}_F$, $\boldsymbol{X}_F$, and $\boldsymbol{M}_F$ for M_F. We also let $\tilde{\boldsymbol{\beta}}_F = (\boldsymbol{X}_F' \boldsymbol{X}_F + \boldsymbol{M}_F)^{-1} \boldsymbol{X}_F' \boldsymbol{y}$, then the full conditional posteriors for $\pi_F(\cdot \mid \boldsymbol{y})$ are given as follows

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

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

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

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

$$B_{RF} = \frac{1}{B_{FR}} \approx \frac{1}{N}\sum_{i=1}^{N}\left|\boldsymbol{I}_n + \sum_{j=R+1}^{R+F} g_j^{(i)}\boldsymbol{P}_j\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2(\sigma^2)^{(i)}}\left(\boldsymbol{y} - \sum_{j=0}^{R}\boldsymbol{X}_j\boldsymbol{\beta}_j^{*(i)}\right)' \left[\boldsymbol{I}_n - \left(\boldsymbol{I}_n + \sum_{j=R+1}^{R+F} g_j^{(i)}\boldsymbol{P}_j\right)^{-1}\right]\left(\boldsymbol{y} - \sum_{j=0}^{R}\boldsymbol{X}_j\boldsymbol{\beta}_j^{*(i)}\right)\right\}.$$

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. Note that only one MCMC step with model M_F is needed for this calculation.

5 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 is discussed. The choice of hyper-parameters in the priors of g is still a problem. In this section, simulation studies are conducted to examine 2 choices of the hyper-parameters. Following the idea of Zellner–Siow prior, b 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 1, 2, 3, 4, 5, and 6. Alternatively, the choice of b can be adjusted according to 'The Effective Sample Size' (Berger et al. 2014) of the corresponding parameters, which is denoted as 'TESS'. For comparison, the generalized Jeffreys–Zellner–Siow prior proposed by Bayarri and 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'.

In the simulation study, we calculate the Bayes factors under different priors to test for the interaction effects in 2-way ANOVA models. Specifically, M_3 and M_4 in Example 1 are considered. Data are 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 1 for an explanation). $\boldsymbol{\beta}_1$ is generated from $N(\boldsymbol{0}, \tilde{g}_1\sigma^2\boldsymbol{I}_{p_1})$, $\boldsymbol{\beta}_2$ is generated from $N(\boldsymbol{0}, \tilde{g}_2\sigma^2\boldsymbol{I}_{p_2})$, $\boldsymbol{\beta}_3$ is generated from $N(\boldsymbol{0}, \tilde{g}_3\sigma^2\boldsymbol{I}_{p_1p_2})$, and $\boldsymbol{\epsilon}$ is generated from $N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{p_1p_2k})$, 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 $\boldsymbol{y}$ are 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 1, 2, 3, 4, 5, and 6.

Table 1 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 2 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 Mean and standard deviation of $\ln(B_{43})$: $k = 5$, $\tilde{g}_3 = 0.1$

Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$				
BG	−12.88 (4.52)	−32.25 (7.58)	−77.7 (10.82)	−273.98 (20.44)
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)
$p_2 = 10$				
BG		−84.85 (10.36)	−228.66 (17.78)	−761.14 (42.78)
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)
$p_2 = 20$				
BG			−591.57 (34.66)	−1936.44 (70.54)
IZS			−114.58 (18.77)	−297.68 (27.71)
TESS			13.14 (6.52)	36.03 (10.26)
$p_2 = 50$				
BG				−6121.23 (169.81)
IZS				−745.49 (46.43)
TESS				96.02 (16.92)

Table 4 Mean and standard deviation of $\ln(B_{43})$: $k = 20$, $\tilde{g}_3 = 0.1$

Method	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 5$				
BG	−11.85 (8.13)	−28.91 (12.94)	−80.88 (17.33)	−272.56 (35.22)
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)
$p_2 = 10$				
BG		−86.47 (17.98)	−218.3 (28.82)	−766.51 (58.7)
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)
$p_2 = 20$				
BG			−596.66 (44.3)	−1951.88 (85.57)
IZS			−46.74 (40.61)	−125.54 (60.99)
TESS			155.04 (26.54)	405.85 (45.33)
$p_2 = 50$				
BG				−6126.18 (232.68)
IZS				−321.85 (103.81)
TESS				1069.93 (73.58)

Table 5 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 6 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)

Lemma 1 *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. □

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 toward M_4 in the sense that sometimes it cannot identify the data from M_3, whereas 'IZS' and 'BG' are biased toward M_3 as they cannot detect the interaction effect when it is weak. We also consider model selection consistency, which means that assuming one of the models under comparison is true, then this true model will be selected if enough data is observed. Discussions can be found in Berger and Pericchi (2001) and Bayarri et al. (2012). 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.

We also conduct another simulation study which tests the main effects in a 2-way ANOVA model. Similarly, we observe that 'BG' and 'IZS' do not always lead to consistent Bayes factor, whereas 'TESS' is consistent. The results are presented in a supplementary document to this article.

6 Real data analyses

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 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.

Table 7 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 8 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

The multiplicity adjusting prior presented in Scott and Berger (2010) is a possible alternative but is not applied here since the focus of this paper is to discuss the effect of the Bayes factor.

6.1 Hald data

The first example is the Hald data, which are 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 7 includes a description of these variables. The Hald data have been commonly used in literature on model selection. See, for example, George and McCulloch (1993); Hald (1952) and Deltell (2011). For both methods being considered, the posterior inclusion probabilities of the covariates are also listed in Table 7.

In Table 8, 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 and 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.

Table 9 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

6.2 Crime data

The second data set includes the crime-related and demographic statistics for 47 US states in 1960 (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 data set ‘UScrime’ in the ‘MASS’ library of R. The response variable is the crime rate, 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 9.

Following other literature, logarithm are taken on all the variables except the indicator variable ‘So’. Table 9 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.

7 Comments

In this paper, 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. The prior is, admittedly, restrictive in terms of the independence structure between different groups of coefficients. One possible solution is to incorporate dependence between different $\boldsymbol{\beta}_j$ as in the original g-prior with corresponding g_j. Another possible way is to decompose the design matrix as Maruyama and George (2011) but incorporating the grouping structure. These warrant our future research. 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. 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.

Appendix A: Proof of Theorem 1

Note that

$$m(\boldsymbol{y} \mid \boldsymbol{g}) = \int_0^\infty \frac{1}{\sigma^2} \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{j=1}^m \left[\frac{|\boldsymbol{X}_j'\boldsymbol{X}_j|^{1/2}}{(2\pi g_j \sigma^2)^{p_j/2}} \right] \int \exp\left\{ -\frac{1}{2\sigma^2}[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\beta}'\boldsymbol{M}\boldsymbol{\beta}] \right\} \mathrm{d}\boldsymbol{\beta}\mathrm{d}\sigma^2.$$

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

$$(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\beta}'\boldsymbol{M}\boldsymbol{\beta} = (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M})(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}) + \boldsymbol{y}'\boldsymbol{R}\boldsymbol{y},$$

where $\boldsymbol{R}$ is defined by (11). Therefore,

$$m(\boldsymbol{y} \mid \boldsymbol{g}) = \int_0^\infty \frac{\exp\left(-\frac{1}{2\sigma^2}\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y}\right)}{(2\pi)^{\frac{n-p_0}{2}}(\sigma^2)^{\frac{n-p_0}{2}+1}|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}|^{\frac{1}{2}}}\prod_{j=1}^m\left(\frac{|\boldsymbol{X}_j'\boldsymbol{X}_j|^{1/2}}{g_j^{p_j/2}}\right)\mathrm{d}(\sigma^2)$$

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

This proves (10).

Appendix B: Proof of Theorem 2

To prove Theorem 2, we first derive some of the necessary results in the following lemma.

Lemma 2 *Suppose that* (13) *holds.*

(a) *Both* $\boldsymbol{P}_{\boldsymbol{\gamma}}$ *in* (14) *and* $\boldsymbol{A}_{\boldsymbol{\gamma}}$ *in* (15) *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,*

$$\left|\boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\right| = \prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\left(1+\sum_{j\in\boldsymbol{\gamma}} g_j\right)^{p_{\boldsymbol{\gamma}}}. \tag{22}$$

(d) *We have the expression for the inverse,*

$$\left[\boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\right]^{-1} = \boldsymbol{I}_n + \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}, \tag{23}$$

where $u_{\boldsymbol{\gamma}}$ *is defined as in* (18).
(e) $u_{\boldsymbol{\gamma}}$ *defined in* (18) *satisfies the following property: for any* $\boldsymbol{\gamma}_0 \in \boldsymbol{\Gamma}$,

$$\sum_{\emptyset\neq\boldsymbol{\gamma}\subseteq\boldsymbol{\gamma}_0} u_{\boldsymbol{\gamma}} = -1 + \frac{1}{1+\sum_{j\in\boldsymbol{\gamma}_0} g_j}. \tag{24}$$

Proof Parts (a) and (b) are easy. For Part (c), the identity matrix $\boldsymbol{I}_n$ can be decomposed as $\boldsymbol{I}_n = \sum_{\boldsymbol{\gamma}\subseteq\{0,\ldots,m\}} \boldsymbol{A}_{\boldsymbol{\gamma}}$. Since $(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j = \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}: j\in\boldsymbol{\gamma}} \boldsymbol{A}_{\boldsymbol{\gamma}}$, we know that

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

$\forall \boldsymbol{\gamma} \subseteq \{0, \ldots, m\}$, $\boldsymbol{A}_{\boldsymbol{\gamma}}$ 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 $\left|\boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j\right| = |\boldsymbol{C}\boldsymbol{C}'| = |\boldsymbol{C}'\boldsymbol{C}|$, and (22) 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

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

Consider the product of $(\boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j)$ and $(\boldsymbol{I}_n + \sum_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}})$, the coefficient before each term $(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}$ should be zero. We use mathematical induction to prove (23). For $(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\{j\}}$, we have $g_j + u_{\{j\}} + g_j u_{\{j\}} = 0$. This implies that $u_{\{j\}} = -g_j/(1 + g_j)$, so (18) holds when $|\boldsymbol{\gamma}| = 1$. For $(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}$ with $|\boldsymbol{\gamma}| = k \geq 2$, if (18) holds for $|\boldsymbol{\gamma}| = k - 1$, we have $u_{\boldsymbol{\gamma}} + \sum_{j \in \boldsymbol{\gamma}} g_j(u_{\boldsymbol{\gamma}} + u_{\boldsymbol{\gamma} \setminus \{j\}}) = 0$, which implies that

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

The conclusion (18) also holds for $|\boldsymbol{\gamma}| = k$. Thus, (23) and (18) are proved.

For Part (e), without loss of generality, we only prove (24) for $\boldsymbol{\gamma}_0 = \{1, \ldots, k\}$. In fact,

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

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. □

Now we are ready to prove Theorem 2. To calculate $\boldsymbol{R}$, we write

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{I}_{p_0} & \boldsymbol{0} \\ -(\boldsymbol{X}^*)'\boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1} & \boldsymbol{I}_{p_1+\cdots+p_m} \end{pmatrix}, \tag{25}$$

where $\boldsymbol{X}^*$ is defined in (8). Then $\boldsymbol{R} = \boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{D}'[\boldsymbol{D}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}']^{-1}\boldsymbol{D}\boldsymbol{X}'$. We get $[\boldsymbol{D}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}']^{-1} = \text{diag}((\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1},\ (\boldsymbol{X}^{*'}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^* + \boldsymbol{M}_1)^{-1})$. Define $\tilde{\boldsymbol{X}} = (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^*$, then $\boldsymbol{R} = (\boldsymbol{I}_n - \boldsymbol{P}_0) - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}'$. Use the fact that for invertible matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Delta}$, $(\boldsymbol{\Phi} + \boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\omega}')^{-1} = \boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\boldsymbol{\omega}(\boldsymbol{\Delta}^{-1} + \boldsymbol{\omega}'\boldsymbol{\Phi}^{-1}\boldsymbol{\omega})^{-1}\boldsymbol{\omega}'\boldsymbol{\Phi}^{-1}$, and define $\boldsymbol{O} = \tilde{\boldsymbol{X}}\boldsymbol{M}_1^{-1}\tilde{\boldsymbol{X}}'$, then

$$\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}' = \boldsymbol{O} - \boldsymbol{O}(\boldsymbol{I}_n + \boldsymbol{O})^{-1}\boldsymbol{O} = \boldsymbol{I}_n - (\boldsymbol{I}_n + \boldsymbol{O})^{-1}. \tag{26}$$

Also,

$$\begin{aligned} \boldsymbol{I}_n + \boldsymbol{O} &= \boldsymbol{I}_n + (\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{X}^*\boldsymbol{M}_1^{-1}\boldsymbol{X}^{*'}(\boldsymbol{I}_n - \boldsymbol{P}_0) \\ &= \boldsymbol{I}_n + (\boldsymbol{I}_n - \boldsymbol{P}_0)\left(\sum_{j=1}^m g_j\boldsymbol{P}_j\right)(\boldsymbol{I}_n - \boldsymbol{P}_0) \ = \ \boldsymbol{I}_n + \sum_{j=1}^m g_j(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_j. \end{aligned} \tag{27}$$

Applying (27) and (23) to (26), we get $\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}+\boldsymbol{M}_1)^{-1}\tilde{\boldsymbol{X}}' = -\sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}}(\boldsymbol{I}_n - \boldsymbol{P}_0)\boldsymbol{P}_{\boldsymbol{\gamma}}$. This proves (17). Next, we calculate $|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}|$. For $\boldsymbol{D}$ defined in (25), $|\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{M}| = |\boldsymbol{D}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{M}\boldsymbol{D}'| = |\boldsymbol{X}_0'\boldsymbol{X}_0|\,|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{M}_1|$. 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

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

By Part (c) of Lemma 2,

$$|\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{M}| = \prod_{j=0}^m |\boldsymbol{X}_j'\boldsymbol{X}_j|\left(\prod_{j=1}^m g_j^{-p_j}\right)\prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\left(1 + \sum_{j\in\boldsymbol{\gamma}} g_j\right)^{p_{\boldsymbol{\gamma}}}. \tag{28}$$

The conclusion (16) follows by plugging (28) into (10). The theorem is proved.

Appendix C: Proof of Theorem 3

For each $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, define $\xi_{\boldsymbol{\gamma}} = \xi(\boldsymbol{\gamma}) = \bigcap_{\xi\in\boldsymbol{\gamma}} \xi$. We first show that for any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$,

$$\boldsymbol{A}_{\boldsymbol{\gamma}} \neq \boldsymbol{0} \Rightarrow \boldsymbol{\gamma} = \{\tau : \tau \supseteq \xi(\boldsymbol{\gamma})\}. \tag{29}$$

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 $(\boldsymbol{I}_n - \boldsymbol{P}_\tau)\boldsymbol{P}_{\xi_\gamma} = \boldsymbol{0}_{n\times n}$ in $\boldsymbol{A}_\gamma$. So (29) holds. Therefore, $\forall \boldsymbol{A}_\gamma \neq \boldsymbol{0}$,

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

so $p_\gamma = \sum_{\tau\subseteq\xi_\gamma} (-1)^{|\xi_\gamma|-|\tau|} p_\tau$. In (16),

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

Next, we need to calculate $\boldsymbol{R}$ in this case. In (17), $\boldsymbol{P}_\gamma = \prod_{\xi\in\gamma} \boldsymbol{P}_\xi = \boldsymbol{P}_{\xi_\gamma}$. Therefore, in (17),

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

The theorem will be proved given the following lemma.

Lemma 3 *For nonempty* $\tau \subseteq \{1, 2, \ldots, q\}$, *define* $U'_\tau = \sum_{\gamma\in\Gamma:\xi(\gamma)=\tau} u_\gamma$. *Then we have*

$$\sum_{\tau^*\supseteq\tau} U'_{\tau^*} = -1 + \frac{1}{1 + \sum_{\tau^*\supseteq\tau} g_{\tau^*}}. \tag{30}$$

$$U'_\tau = \begin{cases} \sum_{\tau^*\supseteq\tau} (-1)^{|\tau^*|-|\tau|} \frac{1}{1+\sum_{\tau^{**}\supseteq\tau^*} g_{\tau^{**}}}, & \text{if } \tau \subsetneq \{1,\ldots,q\}, \\ -1 + \frac{1}{1+g_{\{1,\ldots,q\}}}, & \text{if } \tau = \{1,\ldots,q\}. \end{cases} \tag{31}$$

Proof For (30), note that for any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, $\xi(\boldsymbol{\gamma}) \supseteq \tau \Leftrightarrow \boldsymbol{\gamma} \subseteq \{\tau^* : \tau^* \supseteq \tau\}$. Therefore, using Lemma 2(e), we get

$$\sum_{\tau^*\supseteq\tau} U'_{\tau^*} = \sum_{\emptyset\neq\gamma\subseteq\{\tau^*:\tau^*\supseteq\tau\}} u_\gamma = -1 + \frac{1}{1 + \sum_{\tau^*\supseteq\tau} g_{\tau^*}}.$$

For (31), we use mathematical induction. If $\tau = \{1, \ldots, q\}$, (31) is exactly (30). If $\tau = \{1, \ldots, q-1\}$, from (30), we have $U'_{\{1,\ldots,q\}} + U'_{\{1,\ldots,q-1\}} = -1 + 1/(1 + g_{\{1,\ldots,q-1\}} + g_{\{1,\ldots,q\}})$, which implies that

$$U'_{\{1,\ldots,q-1\}} = \frac{1}{1 + g_{\{1,\ldots,q-1\}} + g_{\{1,\ldots,q\}}} - \frac{1}{1 + g_{\{1,\ldots,q\}}}.$$

This proves that (31) holds for τ with $|\tau| = q - 1$. Clearly, (30) implies a recursive formula,

$$U'_\tau = -1 + \frac{1}{1 + \sum_{\tau^* \supseteq \tau} g_{\tau^*}} - \sum_{\tau^* \supsetneq \tau} U'_{\tau^*}.$$

Suppose (31) holds for τ with $|\tau| = k + 1$, then

$$\begin{aligned} U'_{\{1,\ldots,k\}} &= -1 + \frac{1}{1 + \sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} - \sum_{\tau \supsetneq \{1,\ldots,k\}} U'_\tau \\ &= \frac{1}{1 + \sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} - \sum_{\tau \supsetneq \{1,\ldots,k\}} \left(\sum_{\tau^* \supseteq \tau} (-1)^{|\tau^*| - |\tau|} \frac{1}{1 + \sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} \right) \\ &= \frac{1}{1 + \sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} \\ &\quad - \sum_{\tau^* \supsetneq \{1,\ldots,k\}} \left(\frac{1}{1 + \sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} \left(\sum_{\tau : \tau^* \supseteq \tau \supsetneq \{1,\ldots,k\}} (-1)^{|\tau^*| - |\tau|} \right) \right) \\ &= \frac{1}{1 + \sum_{\tau \supseteq \{1,\ldots,k\}} g_\tau} + \sum_{\tau^* \supsetneq \{1,\ldots,k\}} \left(\frac{1}{1 + \sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} (-1)^{|\tau^*| - k} \right) \\ &= \sum_{\tau^* \supseteq \{1,\ldots,k\}} \left(\frac{1}{1 + \sum_{\tau^{**} \supseteq \tau^*} g_{\tau^{**}}} (-1)^{|\tau^*| - k} \right). \end{aligned}$$

Therefore, (31) holds for τ with $|\tau| = k$. Repeat this procedure recursively, we can show that (31) holds for any nonempty $\tau \subsetneq \{1, \ldots, q\}$. The lemma is proved. □

References

Baksalary, J. K. (1987). Algebraic characterizations and statistical implications of the commutativity of orthogonal projectors. In T. Pukkila, S. Puntanen (Eds.), *Proceedings of the second international Tampere conference in statistics* (pp. 113–142, 1–4 June 1987). Tampere: Department of Mathematical Sciences/Statistics, University of Tampere.

Baksalary, J. K., Baksalary, O. M., Szulc, T. (2002). A property of orthogonal projectors. *Linear Algebra and Its Applications*, *354*, 35–39.

Bayarri, M. J., Berger, J. O., Forte, A., García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, *40*(3), 1550–1577.

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. (2014). The effective sample size. *Econometric Reviews*, *33*, 197–217.

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, volume 38 of IMS Lecture Notes—Monograph Series (pp. 135–207). Institute of Mathematical Statistics.

Berger, J. O., Pericchi, L. R., Varshavsky, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhya: The Indian Journal of Statistics Series A60*(3), 307–321.
Deltell, A. F. (2011). Objective Bayes criteria for variable selection. Ph.D. thesis, Universidad de València.
Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Statistics*, *42*, 204–223.
Fernández, C., Ley, E., Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*(2), 381–427.
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.
Gelman, A. (2005). Analysis of variance why it is more important than ever. *The Annals of Statistics*, *33*(1), 1–53.
George, E. I., McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–889.
Guo, R., Speckman, P. L. (2009). Bayes factor consistency in linear models. In: *The 2009 international workshop on objective Bayes methodology, Philadelphia*, 5–9 June 2009.
Hald, A. (1952). *Statistical theory with engineering applications*. New York: Wiley.
Jeffreys, H. (1961). *Theory of probability*. New York: Oxford University Press.
Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.
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.
Park, T., Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*(482), 681–686.
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.
Rao, C. R., Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *Journal of Statistical Planning and Inference*, *3*, 1–17.
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.
Searle, S., Casella, G., McCulloch, C. (1992). *Variance components. Wiley series in probability and statistics*. New York: Wiley.
Sun, D., Speckman, P. L., Liu, F., Rouder, J. N. (2010). One-way ANOVA, fixed effects or random? (unpublished manuscript).
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* (pp. 270–335). National Academy of Sciences.
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.
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* (pp. 233–243). New York/Amsterdam: Elsevier/North-Holland.
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* (pp. 585–603). Valencia: University of Valencia Press.