

# Mixtures of $g$ -Priors for Analysis of Variance Models with a Diverging Number of Parameters

Min Wang\*

**Abstract.** We consider Bayesian approaches for the hypothesis testing problem in the analysis-of-variance (ANOVA) models. With the aid of the singular value decomposition of the centered designed matrix, we reparameterize the ANOVA models with linear constraints for uniqueness into a standard linear regression model without any constraint. We derive the Bayes factors based on mixtures of  $g$ -priors and study their consistency properties with a growing number of parameters. It is shown that two commonly used hyper-priors on  $g$  (the Zellner-Siow prior and the beta-prime prior) yield inconsistent Bayes factors due to the presence of an inconsistency region around the null model. We propose a new class of hyper-priors to avoid this inconsistency problem. Simulation studies on the two-way ANOVA models are conducted to compare the performance of the proposed procedures with that of some existing ones in the literature.

**Keywords:** ANOVA models, Bayes factor, consistency, growing number of parameters, Zellner's  $g$ -prior.

## 1 Introduction

In the field of applied statistics, analysis-of-variance (ANOVA) is a collection of statistical models commonly used to test hypotheses about the presence of a group (treatment) effect. It has been widely recognized as an important tool to formulate evidence favoring certain theoretical positions and disfavoring others in various areas of application, such as agriculture (VanLeeuwen, 1997), biology (Lazic, 2008), ecology (Qian and Shen, 2007), and psychological studies (Rouder et al., 2012).

We deal with applications of hypothesis testing in the multi-way ANOVA designs, which have been employed by researchers to assess main effects and their interactions in experimental designs. Suppose that  $\mathbf{Y} = [y_1, \dots, y_n]'$  is a random sample of size  $n$  drawn from normally distributed with mean vector  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_n]'$  and covariance matrix  $\sigma^2 \mathbf{I}_n$ , where  $\boldsymbol{\mu}$  and  $\sigma^2$  are both unknown, and  $\mathbf{I}_n$  is an  $n \times n$  identity matrix. The corresponding model can be written as

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n), \quad (1)$$

where  $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the multivariate normal distribution of dimension  $n$  with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The cell mean  $\boldsymbol{\mu}$  can be further decomposed as  $\boldsymbol{\mu} = \mathbf{1}_n \alpha + \mathbf{X} \boldsymbol{\beta}$ , where  $\mathbf{1}_n$  is an  $n \times 1$  vector of ones,  $\mathbf{X}$  represents an  $n \times p$  design

---

\*Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, USA, email: [minwang@mtu.edu](mailto:minwang@mtu.edu)

matrix,  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector of unknown regression coefficients. The model can be reexpressed as

$$\mathbf{Y} \sim N_n(\mathbf{1}_n\alpha + \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n). \quad (2)$$

The design matrix  $\mathbf{X}$  is often referred to as factors in the experiment and is populated by entries of one or zero that describe how effect parameters map onto observations. The regression coefficient  $\boldsymbol{\beta}$  can be viewed as level-specific parameters. Draper and Smith (1998) developed a dummy coding, denoted by `model.matrix(.)` in R language (R Development Core Team, 2011) to construct  $\mathbf{X}$ . Model (2) may not be identifiable, because there are a total of  $p + 1$  parameters that determine the  $p$  cell means, and thus some linear constraints are usually imposed for uniqueness. We here consider the sum-to-zero linear constraints proposed by Fujikoshi (1993). As a result, the intercept becomes the grand mean and each regression coefficient represents the deviation from the grand mean. The regression coefficient of the last level is equal to minus the sum of other regression coefficients. Later on, we will justify that any linear constraint can be adopted without affecting the main results in this paper.

In the ANOVA models, we are interested in the model selection problem where we would like to select a model which is compatible with the observable data. This problem is equivalent to selecting a submodel of (2), whose mean is of the form

$$\boldsymbol{\mu} = \mathbf{1}_n\alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma, \quad (3)$$

where  $\mathbf{X}_\gamma$  is an  $n \times k$  submatrix of  $\mathbf{X}$  and  $\boldsymbol{\beta}_\gamma$  is a  $k \times 1$  vector of unknown regression coefficients. Various procedures have been proposed for the above problem ranging from frequentist ones such as the  $p$ -values and the Akaike information criterion (AIC) to Bayesian methods. From a frequentist viewpoint, researchers routinely report the  $p$ -values as a measure of evidence for competing positions, even though a number of critiques of using the  $p$ -values have been raised in the literature; see Rouder et al. (2012) for detailed comments on the topic. Recently, a growing chorus of researchers advocate the use of the Bayesian procedure as evidence; see, for example, Maruyama (2012), Rouder et al. (2012), Wetzels et al. (2012), Wang and Sun (2013), among others. There are many advantages of using Bayesian approaches for making inference over the frequentist one; see Berger and Pericchi (2001) for a detailed discussion.

For the sake of simplicity, we describe the Bayesian formulation for comparing model (3) with the null model ( $\boldsymbol{\mu} = \mathbf{1}_n\alpha$ ) which does not include any of the predictors. This formulation can be easily adjusted for other model comparison problems. Since the Bayesian approach for the problems of model selection and hypothesis testing is conceptually the same (Guo and Speckman, 2009), we consider the hypothesis testing problem of the form

$$M_1 : \boldsymbol{\mu} = \mathbf{1}_n\alpha \quad \text{versus} \quad M_\gamma : \boldsymbol{\mu} = \mathbf{1}_n\alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma. \quad (4)$$

Within a Bayesian framework, a natural way for evaluating the plausibility of the two competing models is to use the Bayes factor (Kass and Raftery, 1995), which is based

on the posterior model probabilities. The Bayes factor for comparing  $M_\gamma$  to  $M_1$  in (4) can be expressed in terms of the ratio of the two marginal likelihood functions

$$\text{BF}[M_\gamma : M_1] = \frac{m_\gamma(\mathbf{Y})}{m_1(\mathbf{Y})}, \quad (5)$$

where the marginal likelihood of  $\mathbf{Y}$  given  $M_\gamma$  is

$$m_\gamma(\mathbf{Y}) = p(\mathbf{Y} | M_\gamma) = \int \int \int p(\mathbf{Y} | \alpha, \boldsymbol{\beta}_\gamma, \sigma^2) \pi_\gamma(\alpha, \boldsymbol{\beta}_\gamma, \sigma^2) d\alpha d\boldsymbol{\beta}_\gamma d\sigma^2,$$

and the marginal likelihood of  $\mathbf{Y}$  given  $M_1$  is

$$m_1(\mathbf{Y}) = p(\mathbf{Y} | M_1) = \int \int p(\mathbf{Y} | \alpha, \sigma^2) \pi_1(\alpha, \sigma^2) d\alpha d\sigma^2,$$

where  $\pi_\gamma(\alpha, \boldsymbol{\beta}, \sigma^2)$  and  $\pi_1(\alpha, \sigma^2)$  are the joint prior densities for the unknown parameters under  $M_\gamma$  and  $M_1$ , respectively. The Bayes factor for comparing  $M_1$  to  $M_\gamma$ , denoted by  $\text{BF}[M_1 : M_\gamma]$ , is equal to the reciprocal of  $\text{BF}[M_\gamma : M_1]$ . Model  $M_\gamma$  ( $M_1$ ) is more likely to be selected when  $\text{BF}[M_\gamma : M_1] > 1$  ( $< 1$ ). More specifically, Jeffreys (1961) provided the classification categories of evidential strength for the use of the Bayes factor. For example, Jeffreys suggested that  $\text{BF}[M_\gamma : M_1] > 10$ , provides “strong” evidence favoring  $M_\gamma$ , and  $\text{BF}[M_\gamma : M_1] > 100$ , provides “decisive” evidence. Note that the logarithm of the Bayes factor can be interpreted as the weight of evidence provided by the data; see Kass and Raftery (1995).

We need to specify priors for the unknown parameters  $\alpha$ ,  $\boldsymbol{\beta}_\gamma$ , and  $\sigma^2$ . We choose a ‘noninformative’ prior for the common parameters  $\alpha$  and  $\sigma^2$  that appear in both models and place a partially conjugate normal prior on  $\boldsymbol{\beta}_\gamma$  that appears only in the alternative model. We consider Zellner’s  $g$ -prior because it leads to a simple expression of the marginal likelihood. The choice of  $g$  becomes quite crucial in the revival of Bayesian inference because it controls the amount of information in Zellner’s  $g$ -prior. A nice review of mixtures of  $g$ -priors and different fixed values of  $g$  was recently provided by Liang et al. (2008). In this paper, we follow the suggestion of Ley and Steel (2012) and adopt a hyper-prior on  $g$  to reflect its uncertainty and randomness and to allow for the data to determine the inference on  $g$ . This hyper-prior must be proper, because the null model does not involve  $g$  and the improper prior will yield the Bayes factor with an undefined normalizing constant.

Since the seminal work of Zellner and Siow (1980), the Zellner-Siow (ZS) prior has been widely adopted for the unknown parameters in normal linear regression models. Liang et al. (2008) considered three families of hyper-prior for  $g$  (the ZS prior, the hyper- $g$  prior, and the hyper- $g/n$  prior) and studied Bayes factor consistency associated with these priors when the model dimension  $p$  is fixed. Here, consistency means that the true model (hypothesis) will be eventually detected if enough data is provided, assuming that it exists. Later on, Maruyama and George (2011) proposed the beta-prime prior on  $g$ , which yields an analytic Bayes factor without integral representation. They also proved that consistency holds with this prior under the same scenario. Ley and Steel (2012)

conducted Monte Carlo simulations to compare the performance of various mixtures of  $g$ -priors in the literature.

Wetzels et al. (2012) recently generalized mixtures of  $g$ -priors into the ANOVA models, whereas they failed to further examine the performance of the Bayes factors under these priors from both theoretical and practical points of view. Rouder et al. (2012) developed a set of the Bayes factors based on multivariate generalizations of the Cauchy prior and also did not establish consistency property of the considered Bayes factors, even though they emphasized consistency as a desirable theoretical property for a proposed procedure. Bayarri et al. (2012) considered consistency as one of the desired criteria that a model selection prior should satisfy in the context of normal linear regression models, whereas there is not much discussion about this topic in ANOVA designs. This paper fills the gap by studying Bayes factors consistency under various mixtures of  $g$ -priors in the ANOVA models.

Linear models with a growing number of parameters have received considerable attention in the literature. This is often referred to as ‘large  $p$ , large  $n$ ’ regime, in which there is a sizable number of predictors compared to the sample size. Consistency under this regime has been studied in linear models; see, for example, Moreno et al. (2010), Girón et al. (2010), Johnson and Rossell (2012), Wang and Sun (2014), among others. However, the study is still quite scant in ANOVA settings (for exceptions, see Maruyama, 2012 and Wang and Sun, 2013). The question we address here is whether consistency still holds under this regime for the proposed Bayes factors in the ANOVA models. Equivalently, we consider Bayes factor consistency when the model dimension increases with the sample size with rate  $k = O(n^b)$  where  $0 \leq b \leq 1$ . As  $n$  approaches infinity, we focus on the following two asymptotic scenarios

**Scenario 1:**  $k = O(n^b)$  with  $0 \leq b < 1$ .

**Scenario 2:**  $k$  growing proportionally to  $n$ .

We first consider consistency of the Bayes factors under two commonly used hyper priors on  $g$ : the Zellner-Siow prior (Zellner and Siow, 1980) and the beta-prime prior (Maruyama and George, 2011). The first prior is often considered as a default choice for the unknown parameters in the literature; see, for example, Bayarri and García-Donato (2007), Wetzels et al. (2012), Rouder et al. (2012). The second one results in a closed-form Bayes factor, which can thus be calculated as easily as in the case with the fixed values of  $g$ . It deserves mentioning that the proposed results based on the second prior generalize some existing ones for the one-way/two-way ANOVA models studied by Maruyama (2012) and Wang and Sun (2013).

We then show that under very general conditions, the Bayes factors under the above two priors are consistent under the null model and are inconsistent under the alternative model when  $k$  grows proportionally to  $n$ . Although we can explicitly characterize the inconsistency regions in terms of a pseudo-distance between the two competing models, the inconsistency regions could lead to the rejection of the alternative hypothesis when it is true. Of particular note is that the inconsistency region under the Zellner-Siow prior is slightly larger than the one under the beta-prime prior. This finding may justify that the latter outperforms the former from a theoretical viewpoint.

To avoid the inconsistency regions mentioned above, we propose a new family of hyper-priors on  $g$ , which is very flexible and requires little or no a priori input. The Bayes factor based on the proposed prior is not only a closed-form Bayes factor with the unidimensional integral, but also is consistent whichever the true model is and not vulnerable to the Jeffreys-Lindley paradox and the information paradox. More importantly, we demonstrate that the proposed results in the ANOVA models are also valid for the hypothesis testing problem in linear models with a growing number of parameters. The study is of the utmost importance to researchers from both theoretical and practical viewpoints. From a theoretical perspective, it justifies the asymptotic behavior of the proposed procedures for choosing the true model when it exists. From a practical perspective, it provides a guideline to choose appropriate mixtures of  $g$ -priors in many practical applications.

The remainder of the paper is organized as follows. In Section 3, we use the two-way ANOVA model to illustrate how the singular value decomposition (SVD) can be implemented to convert the ANOVA model with constraints into a linear regression model without constraint. In Section 3, we specify priors for the unknown parameters and derive the Bayes factors based on various mixtures of  $g$ -priors. In Section 4, we investigate Bayes factor consistency under various mixtures of  $g$ -priors when  $k = O(n^b)$  where  $0 \leq b \leq 1$ . In Section 5, simulation studies are conducted to evaluate the performance of the considered priors. Finally, some concluding remarks are presented in Section 6, with additional material and proofs given in the Supplementary Material (Wang, 2016).

## 2 Model reparameterization

In this section, we discuss the general development of Bayesian analysis on multi-way ANOVA models. As mentioned in Section 1, we cannot directly implement Zellner's  $g$ -prior for the regression coefficients, because the design matrix (2) does not have full column rank. In order to overcome such difficulty, we implement the sum-to-zero constraints for uniqueness and then consider the SVD of the centralized design matrix. Specifically, we reparameterize the ANOVA model with constraints for uniqueness into a standard linear regression model without any constraint. For the sake of simplicity, we choose the two-way unbalanced ANOVA model to illustrate the implementation of the full-rank model reparameterization.

Consider a factorial design with two treatment factors  $A$  and  $B$  having  $a$  and  $b$  levels, respectively, with a total of  $ab$  factorial cells. Suppose that  $y_{ijk}$  is the  $k$ th observation in the  $(i, j)$ th cell defined by the  $i$ th level of factor  $A$  and the  $j$ th level of factor  $B$ , satisfying the following model

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}, \quad (6)$$

where  $\mu_{ij}$ 's represent the cell mean (expected value), and the residual errors  $\varepsilon_{ijk}$ 's are assumed to be independent random variables, each having a normal distribution with mean zero and unknown variance  $\sigma^2$ . The value of  $n_{ij}$  represents the number of observations in the  $(i, j)$  cell. The total number of observations is  $n = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$ . Model (6) becomes balanced for the case with  $n_{11} = \dots = n_{ab} = m$ .

We decompose the cell mean  $\mu_{ij}$  into the form  $\mu_{ij} = \mu + \alpha_i + \tau_j + \gamma_{ij}$  for  $i = 1, \dots, a$  and  $j = 1, \dots, b$ , where  $\mu$  is the grand mean,  $\alpha_i$  and  $\tau_j$  represent the  $i$ th and  $j$ th main effects of factors  $A$  and  $B$ , respectively, and  $\gamma_{ij}$  is their  $(i, j)$ th interaction. Model (6) can be rewritten as

$$y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n_{ij}. \quad (7)$$

The above model is not identifiable because the parameters  $(\mu, \alpha_i, \tau_j, \gamma_{ij})$  cannot be uniquely defined. Following the work of Fujikoshi (1993), we impose the following linear constraints

$$\sum_{i=1}^a n_{i\cdot} \alpha_i = 0, \quad \sum_{j=1}^b n_{\cdot j} \tau_j = 0, \quad \sum_{i=1}^a \sum_{j=1}^b n_{ij} \gamma_{ij} = 0, \quad (8)$$

where  $n_{i\cdot} = \sum_{j=1}^b n_{ij}$  and  $n_{\cdot j} = \sum_{i=1}^a n_{ij}$ . It is convenient to use matrix notation. Let  $\mathbf{Y} = [\mathbf{Y}'_1, \mathbf{Y}'_2, \dots, \mathbf{Y}'_a]'$  with  $\mathbf{Y}_i = [Y_{i11}, Y_{i12}, \dots, Y_{in_{ib}}]'$ ,  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_a]'$ ,  $\boldsymbol{\tau} = [\tau_1, \dots, \tau_b]'$ ,  $\boldsymbol{\gamma} = [\gamma_{11}, \dots, \gamma_{ab}]'$ , and  $\boldsymbol{\beta}_F = [\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}']'$ . Model (7) can be rewritten compactly in a matrix form as follows

$$\mathbf{Y} = \mu \mathbf{1}_n + \mathbf{Z}_A \boldsymbol{\alpha} + \mathbf{Z}_B \boldsymbol{\tau} + \mathbf{Z}_C \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (9)$$

where  $\boldsymbol{\varepsilon} = [\varepsilon_{111}, \dots, \varepsilon_{abn_{ab}}]'$  follows the multivariate normal distribution with mean vector  $\mathbf{0}_n$  and covariance matrix  $\sigma^2 \mathbf{I}_n$ . We follow the Searle's notations (Searle et al., 1992, pp. 212-213) and let  $\mathbf{Z}_A$ ,  $\mathbf{Z}_B$  and  $\mathbf{Z}_C$  be matrices of orders  $n \times a$ ,  $n \times b$ , and  $n \times ab$ , respectively. They are given by

$$\mathbf{Z}_A = \{d \mathbf{1}_{n_{i\cdot}}\}_{i=1}^a, \quad \mathbf{Z}_B = \{c \{d \mathbf{1}_{n_{ij}}\}_{j=1}^b\}_{i=1}^a, \quad \mathbf{Z}_C = \{d \{d \mathbf{1}_{n_{ij}}\}_{j=1}^b\}_{i=1}^a,$$

where the use of  $c$  and  $d$  within the braces represents that the corresponding partitioned matrices are of the column and diagonal types, respectively. By using the following useful products (Searle et al., 1992)

$$\mathbf{Z}'_A \mathbf{1}_n = \{c \ n_{i\cdot}\}, \quad \mathbf{Z}'_B \mathbf{1}_n = \{c \ n_{\cdot j}\}, \quad \mathbf{Z}'_C \mathbf{1}_n = \{c \ n_{ij}\},$$

it can be readily verified that

$$(\mathbf{X}_F - [n_{1\cdot}, \dots, n_{a\cdot}, n_{\cdot 1}, \dots, n_{\cdot b}, n_{11}, \dots, n_{1b}, n_{21}, \dots, n_{ab}] \otimes \mathbf{1}_n/n)' \mathbf{1}_{a+b+ab} = \mathbf{0}_{a+b+ab},$$

where  $\mathbf{X}_F = [\mathbf{Z}_A, \mathbf{Z}_B, \mathbf{Z}_C]$  is an  $n \times (a + b + ab)$  matrix and  $\otimes$  stands for the Kronecker product. We thus treat  $\mathbf{X}_F - [n_{1\cdot}, \dots, n_{a\cdot}, n_{\cdot 1}, \dots, n_{\cdot b}, n_{11}, \dots, n_{1b}, n_{21}, \dots, n_{ab}] \otimes \mathbf{1}_n/n$  as the centered matrix of  $\mathbf{X}_F$ . We consider the SVD of this centered matrix, namely,

$$\mathbf{X}_F - [n_{1\cdot}, \dots, n_{a\cdot}, n_{\cdot 1}, \dots, n_{\cdot b}, n_{11}, \dots, n_{1b}, n_{21}, \dots, n_{ab}] \otimes \mathbf{1}_n/n = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}',$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are  $n \times (ab - 1)$  and  $(a + b + ab) \times (ab - 1)$  orthogonal matrices, respectively, and  $\boldsymbol{\Sigma}$  is an  $(ab - 1) \times (ab - 1)$  diagonal matrix with positive diagonal

entries. Define  $\boldsymbol{\beta}_F^* = \boldsymbol{\Sigma}\mathbf{V}'\boldsymbol{\beta}_F \in \mathcal{R}^{ab-1}$ . Model (9) is equivalent to the standard linear regression model with unconstrained regression coefficient  $\boldsymbol{\beta}_F^*$  given by

$$\mathbf{Y} = \mu\mathbf{1}_n + \mathbf{U}\boldsymbol{\beta}_F^* + \boldsymbol{\varepsilon}.$$

In the two-way ANOVA model, we are usually concerned about the following five realistic models that include main effects and interactions

$M_1$ : No effect of factor  $A$  and no effect of factor  $B$ , i.e.,  $\boldsymbol{\alpha} = \mathbf{0}_a$ ,  $\boldsymbol{\tau} = \mathbf{0}_b$ ,  $\boldsymbol{\gamma} = \mathbf{0}_{ab}$ .

$M_2$ : Only effect of factor  $A$ , i.e.,  $\boldsymbol{\tau} = \mathbf{0}_b$ ,  $\boldsymbol{\gamma} = \mathbf{0}_{ab}$ .

$M_3$ : Only effect of factor  $B$ , i.e.,  $\boldsymbol{\alpha} = \mathbf{0}_a$ ,  $\boldsymbol{\gamma} = \mathbf{0}_{ab}$ .

$M_4$ : The additive model (without interaction), i.e.,  $\boldsymbol{\gamma} = \mathbf{0}_{ab}$ .

$M_5$ : The full model (with interaction).

With the help of the above full-rank parametrization, the resulting model  $M_\gamma$  can be rewritten as a standard linear model given by

$$M_\gamma : \mathbf{Y} = \mu\mathbf{1}_n + \mathbf{U}\boldsymbol{\beta}_\gamma^*,$$

where  $\mathbf{U}$  is an  $n \times k$  orthogonal matrix,  $\boldsymbol{\beta}_\gamma^* \in R^k$  for  $\gamma = 2, \dots, 5$ , and

$$k = \begin{cases} a - 1, & \text{under } M_2, \\ b - 1, & \text{under } M_3, \\ a + b - 2, & \text{under } M_4, \\ ab - 1, & \text{under } M_5. \end{cases}$$

We here used the two-way unbalanced ANOVA model to illustrate how the SVD can be implemented to reparameterize the model with constraints into a standard linear regression model without constraint. In a similar way as done above, we can generalize the previous developments to the hypothesis testing problem (4) in multi-way ANOVA models and obtain that

$$M_1 : \mathbf{Y} = \mu\mathbf{1}_n \quad \text{versus} \quad M_\gamma : \mathbf{Y} = \mu\mathbf{1}_n + \mathbf{U}\boldsymbol{\beta}_\gamma^*, \quad (10)$$

where  $\boldsymbol{\beta}_\gamma^*$  is a  $k$ -dimensional vector of regression coefficients and  $\mathbf{U}$  is an  $n \times k$  orthogonal matrix such that  $\mathbf{U}'\mathbf{U} = \mathbf{I}_k$ . It should be noted that the value of  $k$  is determined by the levels of factors in the alternative model under consideration. One advantage of such reparameterization is that it avoids the difficulty of directly implementing Zellner's  $g$ -prior for the alternative model in (4), whose design matrix  $\mathbf{X}_\gamma$  does not have full column rank.

It is remarkable that Bayarri and García-Donato (2007) proposed the Bayes factor based on the Zellner-Siow prior for testing the general hypotheses in normal linear models, which do not require the design matrix to be of full rank. The Bayes factor has a simple expression with the unidimensional integral. They generalized the Bayes factor to a variety of problems in which the null hypotheses are given by general linear

restrictions. In the two-way ANOVA models, we observe that the constraints in (8) can be expressed as a form of  $\mathbf{C}\boldsymbol{\beta}_F = \mathbf{0}$  with

$$\mathbf{C} = \begin{bmatrix} n_{1.} & \dots & n_{a.} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & n_{.1} & \dots & n_{.b} & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & n_{11} & \dots & n_{ab} \end{bmatrix}. \quad (11)$$

This observation shows that the model reparameterization based on the SVD in this paper is equivalent to the full-rank factorization technique described in Proposition 5 of Bayarri and García-Donato (2007). Consequently, we may conclude from Theorem 1 of Bayarri and García-Donato (2007) that any linear constraint can be adopted for uniqueness without affecting our Bayes factors and their corresponding theoretical properties in this paper.

### 3 Mixtures of $g$ -priors

Bayesian analysis begins with prior specifications for the unknown model parameters. For our hypothesis testing problem (10), we need to specify priors for the unknown parameters  $\alpha$ ,  $\sigma^2$ , and  $\boldsymbol{\beta}_\gamma^*$ . We can regard  $\alpha$  and  $\sigma^2$  as ‘common’ parameters that appear in both models  $M_1$  and  $M_\gamma$  and thus specify the right-Haar prior

$$p(\alpha, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (12)$$

One may refer to Berger and Pericchi (1996) for an asymptotic justification of the use of the same (even noninformative) prior on the common parameters. A more solid justification has recently been provided by Bayarri et al. (2012) based on the group invariance and predictive matching arguments. Since  $\mathbf{U}'\mathbf{U} = \mathbf{I}_k$ , we assign Zellner’s  $g$ -prior on  $\boldsymbol{\beta}_\gamma^*$

$$p(\boldsymbol{\beta}_\gamma^* | \sigma^2, g) \sim N(\mathbf{0}_k, g\sigma^2\mathbf{I}_k),$$

where  $\mathbf{0}_k$  is a  $k \times 1$  vector of zeros. The scaling factor  $g$  controls the amount of information of the prior, so its choice is very critical in the revival of Bayesian inference and will be discussed later. The resulting Bayes factor for comparing  $M_\gamma$  and  $M_1$  is given by

$$\text{BF}[M_\gamma : M_1](g) = (1 + g)^{(n-k-1)/2} (1 + g(1 - R^2))^{-(n-1)/2}, \quad (13)$$

where  $R^2$  is the coefficient of determination of  $M_\gamma$  in (10).

There has been a large fraction of the literature dealing with the choice of  $g$ , which can be categorized into two types: a fixed value of  $g$  and a hyper-prior on  $g$ . George and Foster (2000) commented that fixed values of  $g$  may cause some undesirable behavior: large values would favor the null model, while small values result in the prior dominating the likelihood. We thus choose a proper hyper-prior on  $g$ , denoted by  $\pi(g)$ , which yields the Bayes factor

$$\text{BF}[M_\gamma : M_1] = \int_0^\infty (1 + g)^{(n-k-1)/2} (1 + g(1 - R^2))^{-(n-1)/2} \pi(g) dg. \quad (14)$$



We observe that a number of hyper-priors on  $g$  are obtained in the literature by specifying priors on the shrinkage factor  $g/(1+g)$ , mainly because their properties are often evaluated in terms of this shrinkage factor; see Ley and Steel (2012). In the following section, we consider two commonly used hyper-priors on  $g$  and propose a new type of hyper-priors to avoid the inconsistency issue encountered by the existing procedures.

### 3.1 The Zellner-Siow prior

Inspired by Jeffreys (1961) comments of using the Cauchy prior for comparing a univariate normal mean, Zellner (1986) proposed the multivariate Cauchy distribution for the regression coefficients in normal linear models. This prior is equivalent to Zellner's  $g$ -prior with an inverse gamma distribution on  $g$

$$\pi^{\text{ZS}}(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} \exp\left(-\frac{n}{2g}\right),$$

which corresponds to the following prior for the shrinkage factor  $\vartheta = g/(1+g)$

$$\pi(\vartheta) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} \vartheta^{-3/2} (1-\vartheta)^{-1/2} \exp\left(-\frac{n(1-\vartheta)}{2\vartheta}\right).$$

The Zellner-Siow prior is often adopted to derive 'default' Bayes factor approaches in the Bayesian literature; see, for example, Wetzels et al. (2012), Rouder et al. (2012). The Bayes factor under this prior has a simple expression

$$\text{BF}^{\text{ZS}}[M_\gamma : M_1] = \frac{(n/2)^{1/2}}{\Gamma(1/2)} \int_0^\infty (1+g)^{(n-k-1)/2} [1+g(1-R^2)]^{-(n-1)/2} g^{-3/2} e^{-n/(2g)} dg, \quad (15)$$

which can be easily calculated using unidimensional integration technique. Alternatively, we may employ Monte Carlo approximation by generating samples from an inverse gamma distribution (Bayarri and García-Donato, 2007) or the Laplace approximation with a change of variables (Liang et al., 2008).

### 3.2 The beta-prime prior

Maruyama and George (2011) recently assigned the beta-prime prior on  $g$

$$\pi^{\text{BP}}(g) = \frac{g^t (1+g)^{-s-t-2}}{B(s+1, t+1)}, \quad s > -1, \quad t > -1,$$

which is obtained by specifying a beta distribution with parameters  $s+1$  and  $t+1$  for  $1-\vartheta$ . Later on, Maruyama (2012) derived several expressions of the Bayes factors based on this prior in the one-way/two-way ANOVA models. In this paper, we propose a unified expression of these Bayes factors, which can be implemented the multi-way ANOVA models. Simple algebra shows that the Bayes factor can be unified as

$$\text{BF}^{\text{BP}}[M_\gamma : M_1] = \frac{1}{B(s+1, t+1)} {}_2F_1\left(\frac{n-1}{2}, t+1; \frac{k}{2} + s + t + 2, R^2\right),$$

where  ${}_2F_1$  is the Gaussian hypergeometric function defined as

$${}_2F_1(\alpha, \beta; \gamma, z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha} dt \quad (16)$$

with  $\gamma > \beta > 0$ . The  ${}_2F_1$  in (16) can be numerically estimated by using subroutines in the Cephes library (<http://www.netlib.org/cephes>). With the particular choice of  $t$ , we derive an explicit Bayes factor summarized in the following theorem with its proof in Appendix A of the Supplementary Material.

**Theorem 1.** *Under the beta-prime prior with  $t = (n - k - 1)/2 - s - 2$ , the Bayes factor (14) for comparing  $M_\gamma$  and  $M_1$  in (10) is given by*

$$\text{BF}^{\text{BP}}[M_\gamma : M_1] = \frac{\Gamma(k/2 + s + 1)\Gamma((n - k - 1)/2)}{\Gamma(s + 1)\Gamma((n - 1)/2)} (1 - R^2)^{-(n-k-3)/2+s}. \quad (17)$$

Such a closed-form expression is unavailable for other choices of  $t$ . By following the suggestions by Maruyama and George (2011), we recommend the choice of  $s \in (-1, -1/2]$  for practical applications. Of particulate note is that when the number of parameters under  $M_\gamma$  is bounded, they are asymptotically equivalent to the Bayesian information criterion (BIC) summarized in the following theorem.

**Theorem 2.** *When  $k$  is fixed and  $n$  is sufficiently large, the Bayes factors in (15) and (17) can be approximated by*

$$\text{BF}^{\text{ZS}}[M_\gamma : M_1] \approx \text{BF}^{\text{BP}}[M_\gamma : M_1] \approx \left(\frac{n}{2}\right)^{-k/2} (1 - R^2)^{-(n-k)/2}. \quad (18)$$

*Proof.* The proof directly follows Stirling's formula for the gamma function (in the Supplementary Material) and some algebraic manipulations.  $\square$

**Remark 1.** In the context of linear regression models, Moreno et al. (2015) also showed that the intrinsic Bayes factor is asymptotically equivalent to the BIC, illustrating that the three Bayes factors have the same asymptotic behavior with the BIC when  $k$  is bounded and  $n$  is sufficiently large.

### 3.3 The proposed prior

In the  $g$ -prior framework, Ley and Steel (2012) commented that as the sample size  $n$  increases, we should account for the fact that the information accrues with the sample size and that the inverse of the information matrix is of order  $1/n$ . They suggested the choice of prior distribution on  $g/n$ , instead of  $g$ . Equivalently, we should choose the beta distribution on the shrinkage factor  $v = g/(n + g)$ , This suggestion results in the hyper- $g/n$  prior given by

$$\pi^{\text{HGn}}(g) = \frac{a-2}{2n} \left(1 + \frac{g}{n}\right)^{-a/2}, a > 2, \quad (19)$$

which can be viewed as a modification of the hyper- $g$  prior ( $\pi^{\text{HG}}(g) = (a-2)/2(1+g)^{a/2}$ ); see Liang et al. (2008) in detail about the use of the hyper- $g/n$  prior.

When the model dimension  $k$  grows with  $n$ , it seems natural to take the accrual of the information with both  $n$  and  $k$  into consideration. A possible way is to specify prior distributions on  $g/r$  rather than  $g$  or  $g/n$ , where  $r = n/k$ . This motivates us to choose a beta distribution with parameters  $\alpha + 1$  and  $\beta + 1$  for the modified shrinkage factor  $\varphi = g/(r + g)$ , instead of  $v = g/(n + g)$ . The prior is given by

$$\pi(\varphi) = \frac{\varphi^\beta(1 - \varphi)^\alpha}{B(\alpha + 1, \beta + 1)}, \quad \alpha > -1, \quad \beta > -1,$$

which leads to the following hyper-prior on  $g$

$$\pi(g) = \frac{1}{rB(\alpha + 1, \beta + 1)} \left(\frac{g}{r}\right)^\beta \left(1 + \frac{g}{r}\right)^{-\alpha-\beta-2}, \quad (20)$$

The prior in (20) is the Pearson type VI distribution (Pearson, 1895) with shape parameters  $\alpha > -1$ ,  $\beta > -1$ , and scale parameter  $r > 0$ . This prior has a density whose right tail behaves like  $(g/r)^{-(\alpha+2)}$ , thus providing very fat tails for small values of  $\alpha$ . We suggest the choice of  $\alpha \in (-1, -1/2]$ . The Bayes factor under this prior has a simple expression with the unidimensional integral and can thus be easily computed by numerical integration techniques. Alternatively, we can use the Monte Carlo method by generating  $t_1 \cdots, t_N$  observations from the Pearson type VI distribution with the parameters  $\alpha$ ,  $\beta$ , and  $r$ . The Bayes factor can be approximated by

$$\frac{1}{N} \sum_{i=1}^N (1 + t_i)^{(n-k-1)/2} (1 + t_i(1 - R^2))^{-(n-1)/2}.$$

Note that the Pearson type VI random variables can be easily generated by using `rpearsonVI()` in the R-package of *PearsonDS*; see Becker and Klößner (2013).

We consider several choices of the hyperparameters to illustrate the flexibility of the proposed prior in (20). We start from the hyper- $g/r$  prior with  $\alpha = a/2 - 2 > -1$  (i.e.,  $a > 2$ ) and  $\beta = 0$

$$\pi^{\text{HG}r1}(g) = \frac{a-2}{2r} \left(1 + \frac{g}{r}\right)^{-a/2}, \quad (21)$$

which is similar to the hyper- $g/n$  prior in (19) proposed by Liang et al. (2008). The main difference between the two priors is that the prior in (21) depends on  $k$  only through  $r$ . This is a key feature to study the asymptotic behavior of the Bayes factors under the scenario in which  $k = O(n^b)$  with  $0 \leq b \leq 1$ .

The proposed prior in (20) corresponds to the horseshoe prior (Carvalho et al., 2010) developed in a different setting if we choose  $\alpha = \beta = -1/2$ . It is therefore of interest to study the effects of the shrinkage factor  $\varphi$  if we have the U-shape prior with its spike around 0, which provides very strong shrinkage and thus induces zero regression

coefficients in the shrinkage-prior framework. We thus consider the proposed prior with  $\alpha = \beta = -1/2$

$$\pi^{\text{HG}r2}(g) = \frac{1}{2\sqrt{rg}} \left(1 + \frac{g}{r}\right)^{-3/2}. \quad (22)$$

The proposed prior in (20) is also related to a natural ‘generic’ prior distribution for the shrinkage factor  $\varphi$ , which specifies a uniform prior on the interval  $[0, 1]$  if we choose  $\alpha = \beta = 0$ . The prior is given by

$$\pi^{\text{HG}r3}(g) = \frac{1}{r} \left(1 + \frac{g}{r}\right)^{-2}. \quad (23)$$

The choice of the uniform prior on  $\varphi$  is often treated as reference purposes of Bayesian analysis in the literature. Finally, we propose a general family of hyper-priors on  $g$  given by

$$\pi^{\text{HG}r}(g) = \frac{1}{2\sqrt{rg}} f_r \left( \frac{\sqrt{g}}{\sqrt{r}} \right), \quad (24)$$

where  $f_r(\cdot)$  is a function satisfying  $\int_0^\infty f_r(t) dt = 1$ . The Bayes factor under the prior in (24) is given by

$$\text{BF}^{\text{HG}r}[M_\gamma : M_1] = \int_0^\infty (1 + rt^2)^{(n-k-1)/2} [1 + rt^2(1 - R^2)]^{-(n-1)/2} f_r(t) dt. \quad (25)$$

It should be noted that the Bayes factor under the prior in (20) is just a special case of the one in (25) if we choose  $f_r(t)$  as

$$f_r(t) = \frac{2t^{2\beta+1}(1+t^2)^{-\alpha-\beta-2}}{B(\alpha+1, \beta+1)}.$$

We have considered three different types of hyper-priors on  $g$  in the ANOVA models: the Zellner-Siow prior, the beta-prime prior, and the new proposed prior. It is therefore of interest to compare the performance of the Bayes factors under these priors from both theoretical and practical points of view.

## 4 Bayes factor consistency

In this section, we consider the asymptotic behavior of the Bayes factors under various mixtures of  $g$ -priors. We focus on the information paradox in Section 4.1, and the model selection consistency in Section 4.2.

### 4.1 Information paradox of mixtures of $g$ -priors

Suppose, for the hypothesis testing problem in (10), model  $M_\gamma$  accounts for an overwhelming amount of the variability of data compared to model  $M_1$ . In this setting,

with both  $n$  and  $k$  fixed,  $R^2$  should approach 1 (or equivalently, the usual  $F$ -statistic approaches infinity). We anticipate that the Bayes factor in (13) goes to infinity as the information against  $M_1$  accumulates, whereas it converges to a constant  $(1+g)^{(n-k-1)/2}$  for a fixed choice of  $g$ . This phenomenon is often called the information paradox; see, for example, Jeffreys (1961), Wang and Sun (2013), among others.

**Theorem 3.** *With both  $n$  and  $k$  fixed satisfying  $n > k + 1$ , and  $R^2 \rightarrow 1$ , the Bayes factor in (15) avoids the information paradox, and the Bayes factor in (17) also avoids the information paradox when  $-1 < s < (n - k - 3)/2$ .*

*Proof.* The proof of the result for the Bayes factor in (15) follows directly from Theorem 2 of Liang et al. (2008) and is thus omitted for simplicity. When  $R^2 \rightarrow 1$ , the Bayes factor in (17) approaches infinity for  $-(n - k - 3)/2 + s < 0$ , indicating that it avoids the information paradox given that  $-1 < s < (n - k - 3)/2$ .  $\square$

We here present a general condition under which the proposed prior in (24) resolves the information paradox.

**Theorem 4.** *With both  $n$  and  $k$  fixed satisfying  $n > k + 1$ , and  $R^2 \rightarrow 1$ , the Bayes factor in (25) avoids the information paradox whenever  $\int_0^\infty (1 + rt^2)^{(n-k-1)/2} f_r(t) dt = \infty$ .*

*Proof.* The integrand of the Bayes factor in (25) is a monotonic increasing function of  $R^2$ . By using the monotonic convergence theorem, the Bayes factor tends to  $\int_0^\infty (1 + rt^2)^{(n-k-1)/2} f_r(t) dt$  as  $R^2 \rightarrow 1$ . Thus, the nonintegrability of  $(1 + rt^2)^{(n-k-1)/2} f_r(t)$  is a necessary and sufficient condition for resolving the information paradox.  $\square$

It deserves mentioning that the above three different mixtures of  $g$ -priors may fail to resolve the information paradox with a minimal sample size. For instance, the information paradox exists in the one-way ANOVA model with a fixed-effect having 2 levels with 1 observation in each level.

## 4.2 On consistency of mixtures of $g$ -priors

We study Bayes factor consistency associated with various mixtures of  $g$ -priors when  $k = O(n^b)$  for  $0 \leq b \leq 1$ . Consistency means that the true hypothesis will be selected if enough data is provided, assuming that one of the hypotheses is true. According to Fernández et al. (2001), the Bayes factor is said to be consistent if

$$\text{plim}_{n \rightarrow \infty} \text{BF}[M_\gamma : M_1] = \infty,$$

if  $M_\gamma$  is the true model, whereas

$$\text{plim}_{n \rightarrow \infty} \text{BF}[M_\gamma : M_1] = 0,$$

if  $M_1$  is the true model, where ‘plim’ stands for convergence in probability. As  $n \rightarrow \infty$ , the asymptotic behavior of the Bayes factor depends on the pseudo-distance between the two competing models. We define the pseudo-distance from  $M_\gamma$  to  $M_1$  in (10) as

$$\delta_n = \frac{\boldsymbol{\beta}_\gamma^{*\prime} \boldsymbol{\beta}_\gamma^*}{n\sigma^2} = \frac{\boldsymbol{\beta}_\gamma' \mathbf{V} \boldsymbol{\Sigma} \boldsymbol{\Sigma}' \mathbf{V}' \boldsymbol{\beta}_\gamma}{n\sigma^2},$$

where  $\boldsymbol{\beta}_\gamma^*$  is the regression coefficients of the alternative model. For simplicity of notation, we assume that under the alternative model, the limit of the distance exists and is denoted by  $\delta = \lim_{n \rightarrow \infty} \delta_n$ .

**Theorem 5.** *We consider the Bayes factors  $\text{BF}^{\text{BP}}$  in (17) and  $\text{BF}^{\text{ZS}}$  in (15) for the hypothesis testing problem given by (4) in the ANOVA models.*

- (a) *Under Scenario 1:  $k = O(n^b)$  with  $0 \leq b < 1$ , they are consistent whichever the true model is.*
- (b) *Under Scenario 2:  $k$  growing proportionally to  $n$ , they are consistent under  $M_1$ . However,  $\text{BF}^{\text{BP}}$  is consistent under  $M_\gamma$  when  $\delta > \xi(r)$ , where*

$$\xi(r) = r^{1/(r-1)} - 1. \quad (26)$$

*$\text{BF}^{\text{ZS}}$  is consistent under  $M_\gamma$  when  $\delta > Q(r, 1/r - 1/r^2)$  where  $Q(x, \tau)$  with positive  $\tau$  is the (unique) positive solution of*

$$\frac{(1+x)^r}{re(1-1/r)^{r-1}} - (1+x) - \tau > 0. \quad (27)$$

*Proof.* See Appendix B of the Supplementary Material. □

Part (a) of Theorem 5 shows that under Scenario 1, the two Bayes factors asymptotically choose the true model whichever the true model is. However, Part (b) indicates that under Scenario 2, they fail to be consistent under  $M_\gamma$  due to the presence of a small inconsistency region around  $M_1$  and that we can characterize the inconsistency region with the pseudo-distance between the two models. Figure 1 shows that as  $r$  increases, the two regions approach each other and will eventually disappear when  $r$  tends to infinity. In other words, equations (26) and (27) are both decreasing convex functions of  $r$ , satisfying  $\lim_{r \rightarrow \infty} \xi(r) = 0$  and  $\lim_{r \rightarrow \infty} Q(r, \tau) = 0$  with  $\tau$  being the solution of equation (27). We observe from Figure 1 that the inconsistency region of  $\text{BF}^{\text{BP}}$  is narrower than the one of  $\text{BF}^{\text{ZS}}$ . We may thus conclude that  $\text{BF}^{\text{BP}}$  outperforms  $\text{BF}^{\text{ZS}}$  from a theoretical point of view.

In the context of the one-way ANOVA model, Theorem 5 seems to be in contradiction with Theorem 3.2 of Berger et al. (2003), which states that under Scenario 2 with known  $\sigma^2$ , the Bayes factor under the general form of the prior (Equation 14 of Berger et al., 2003) is always consistent under the alternative model if  $\delta > 0$ . Under the same scenario,

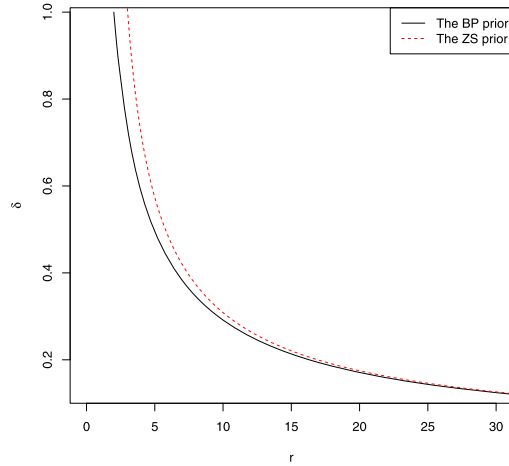


Figure 1: The inconsistency region (an area between the curve and x-axis) for the Bayes factor under Scenario 2 when sampling from  $M_\gamma$ .

Girón et al. (2010) also observed a similar contradictory result for the Bayes factor based on the intrinsic prior, whose inconsistency region is given by  $\delta > \kappa(r)$ , where

$$\kappa(r) = \frac{r - 1}{(r + 1)^{(r-1)/r} - 1} - 1, \tag{28}$$

which is a decreasing function of  $r$  satisfying  $\lim_{r \rightarrow \infty} \kappa(r) = 0$ . The main reason of this contradiction is that the prior of the general form studied by Berger et al. (2003) incorporates the prior with its variance tending to 0, indicating that the mass of the prior is getting less and less to any neighborhood of the null model. Thus, the inconsistency region around the null model will disappear as the variance of the prior approaches 0. We observe that the asymptotic behavior of  $\text{BF}^{\text{ZS}}$  is quite similar to the intrinsic Bayes factor, because the intrinsic prior can be represented as a scaled mixture of  $g$ -priors with a beta mixing distribution; see Lemma 1 of Womack et al. (2014). We here refer the interested reader to Section 4 of Girón et al. (2010) for a further discussion about the inconsistency region of the intrinsic Bayes factor.

**Theorem 6.** *We consider the Bayes factor  $\text{BF}^{\text{HGr}}$  in (25) for the hypothesis testing problem given by (4) in the ANOVA models. Consider a sequence of the proper priors in (24) satisfying two conditions (i)  $\lim_{r \rightarrow \infty} f_r(t)$  exists and (ii)  $\int_0^\infty f_r(t) dt = 1$ .*

- (a) *Under Scenario 1:  $k = O(n^b)$  with  $0 \leq b < 1$ ,  $\text{BF}^{\text{HGr}}$  is consistent whichever model is true.*
- (b) *Under Scenario 2:  $k$  growing proportionally to  $n$ ,  $\text{BF}^{\text{HGr}}$  is consistent whichever model is true.*

*Proof.* See Appendix C of the Supplementary Material. □

This theorem shows that the Bayes factor under the proposed prior in (24) is always consistent whichever the true model is when  $k = O(n^b)$  with  $0 \leq b \leq 1$ . As an illustration, we compare the performance of the two Bayes factors under the various mixtures of  $g$ -priors in the one-way ANOVA model. To mimic the asymptotic scenario, let  $n = 1,000$  and the levels of the treatment factor be 100, indicating that the average number of observations per level is  $r = 10$ . Suppose in particular that we take  $\delta = \{0.2, 0.3, 0.35\}$ . Numerical results are presented in Table 1. As one would expect, the alternative model  $M_\gamma$  is more favorable than the null model  $M_1$  with the large sample size. When  $\delta = 0.2$ ,  $\text{BF}^{\text{BP}}$  and  $\text{BF}^{\text{ZS}}$  are both in favor of  $M_1$ . When  $\delta = 0.3$ ,  $\text{BF}^{\text{BP}}$  provides strong evidence against  $M_1$ , whereas  $\text{BF}^{\text{ZS}}$  is still in favor of  $M_1$  because  $\delta = 0.3 < 0.3086$ . These results are quite reasonable because the two values of  $\delta$  fall in the corresponding inconsistency regions. As expected, when  $\delta = 0.35$ , they both correctly support  $M_\gamma$ . This simple example confirms the results of Theorem 5 and shows that  $\text{BF}^{\text{BP}}$  has a smaller inconsistency region than  $\text{BF}^{\text{ZS}}$  when the model dimension grows proportionally to the sample size  $n$ . In addition, this simple example further confirmed the results of Theorem 6.

Bayes Factor	consistency region	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.35$
$\text{BF}^{\text{BP}}[M_\gamma : M_1]$	$\delta > 0.2915$	-32.80207	2.917074	19.76572
$\text{BF}^{\text{ZS}}[M_\gamma : M_1]$	$\delta > 0.3086$	-37.56758	-1.809035	15.23796
$\text{BF}^{\text{HG}1}[M_\gamma : M_1]$	$\delta > 0$	33.59134	59.00707	71.88346
$\text{BF}^{\text{HG}2}[M_\gamma : M_1]$	$\delta > 0$	32.78827	58.40660	71.35980
$\text{BF}^{\text{HG}3}[M_\gamma : M_1]$	$\delta > 0$	33.38731	58.96554	71.89989

Table 1: Comparisons of the logarithms of the Bayes factors with the various mixtures of  $g$ -priors when  $M_\gamma$  is favorable.

In summary, these Bayes factors are always consistent when sampling from the null model  $M_1$ . Table 2 provides a brief summary of comparisons among these Bayes factors when sampling from the alternative model  $M_\gamma$ . The Bayes factors based on the ZS prior and the BP prior behave similarly, both having an inconsistency region around the null model under Scenario 2, whereas the Bayes factor based on the proposed family of hyper-priors avoids this inconsistency region.

	$\text{BF}^{\text{BP}}[M_\gamma : M_1]$	$\text{BF}^{\text{ZS}}[M_\gamma : M_1]$	$\text{BF}^{\text{HG}r}[M_\gamma : M_1]$
Scenario 1	consistent	consistent	consistent
Scenario 2	$M_\gamma: \delta > \xi(r)$	$M_\gamma: \delta > Q(r, 1/r - 1/r^2)$	consistent

Table 2: Comparison among the Bayes factors under the various hyper-priors on  $g$  when sampling from  $M_\gamma$  under the two asymptotic scenarios.

**Remark 2.** For the problem of variable selection in normal linear models, Liang et al. (2008) adopted a particular form of Zellner’s  $g$ -prior, which results in the Bayes factor for comparing any non-null model  $M_\gamma$  to the null model  $M_1$

$$\text{BF}^{\text{L}}[M_\gamma : M_1] = \int_0^\infty (1 + g)^{(n-p_\gamma-1)/2} (1 + g(1 - R_\gamma^2))^{-(n-1)/2} \pi(g) dg, \quad (29)$$



where  $p_\gamma$  is the number of non-zero regression coefficients and  $R_\gamma^2$  is the ordinary coefficient of determination of model  $M_\gamma$ . By replacing  $k$  and  $R^2$  with  $p_\gamma$  and  $R_\gamma^2$ , respectively, the Bayes factor in (14) in the ANOVA models is the same as the one proposed by Liang et al. (2008) in linear models. This connection shows that the proposed Bayes factors and their attractive properties (Theorems 5 and 6) remain valid in the context of linear regression models under the same asymptotic scenarios if we assume the existence of the pseudo-distance from  $M_\gamma$  to  $M_1$  given by

$$\delta_n = \frac{\boldsymbol{\beta}_\gamma' \mathbf{X}_\gamma' \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma}{n\sigma^2},$$

where  $\boldsymbol{\beta}_\gamma$  is the regression coefficients of model  $M_\gamma$ . Equivalently, when the sample size and the number of parameters approach infinity, the limit of this distance exists and is given by  $\delta = \lim_{n \rightarrow \infty} \delta_n > 0$  when sampling from  $M_\gamma$ .

## 5 Simulation study

In this section, we conduct simulation studies to compare the performance of the Bayes factors under the various mixtures of  $g$ -priors in the two-way ANOVA models. We consider the three special choices of the proposed prior: the HGr1 prior in (21), the HGr2 prior in (22) with  $a = 3$ , and the HGr3 prior in (23). Along with these priors, we also consider two commonly used mixtures of  $g$ -priors: the Zellner-Siow (ZS) prior and the beta-prime (BP) prior with  $s = -1/2$ .

As an illustration, we mainly focus on the model comparison between  $M_5$  and  $M_1$  described in Section 2. The results from other model comparisons are similar and are thus omitted for simplicity. Without loss of generality, let  $\mu = 0$ , and  $\sigma^2 = 1$ ; see Min and Sun (2015). We simulate the data from the sampling models as follows. For the two-way ANOVA model (9) in Section 2,  $\boldsymbol{\alpha}$  is generated from  $N(\mathbf{0}_a, g\sigma^2\mathbf{I}_a)$ ,  $\boldsymbol{\tau}$  from  $N(\mathbf{0}_b, g\sigma^2\mathbf{I}_b)$ ,  $\boldsymbol{\gamma}$  from  $N(\mathbf{0}_{ab}, g\sigma^2\mathbf{I}_{ab})$ , and  $\boldsymbol{\varepsilon}$  from  $N(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$ , where  $g = \{0, 0.05, 0.2\}$ . When  $g = 0$ , the data are from the null model  $M_1$ ; when  $g > 0$ , the data are from the full model  $M_5$ . To mimic the two asymptotic scenarios in the Introduction, we consider two different choices of  $\{m, a, b\}$  in model (6):  $\{a = 2, b = 3, m = 400\}$  and  $\{a = 30, b = 30, m = 6\}$ . It is important to remember that  $k = ab - 1$  and  $r = abm/k$  when comparing  $M_5$  and  $M_1$ . For each case, we replicate our simulation 1,000 times. The decision criterion is to select  $M_5$  if  $\log(\text{BF}) > 0$ , and  $M_1$  otherwise.

The five number summary of  $\log(\text{BF})$  and the relative frequency (RF) of selecting the correct model under each prior are presented in Tables 3 and 4. Rather than providing exhaustive interpretations from these simulations, we merely highlight the most important findings as follows. (i) When  $\{a = 2, b = 3, m = 400\}$ , the performance of the proposed Bayes factors are comparable with the ones under the two commonly used mixtures, because they all provide the desirable results whichever the sampling model is. (ii) When  $\{a = 30, b = 30, m = 6\}$  and  $g = 0.05$ , Table 4 shows that the logarithms of  $\text{BF}^{\text{ZS}}$  and  $\text{BF}^{\text{BP}}$  are all negative and thus fail to provide evidence supporting the true model. However, the proposed Bayes factors are always positive and are in favor of the correct model. (iii) When  $\{a = 30, b = 30, m = 6\}$  and  $g = 0.20$ ,  $\text{BF}^{\text{BP}}$  outperforms

$g$	prior	Min	$Q_1$	Median	$Q_3$	Max	RF(%)
0	ZS	-16.6781	-14.2927	-12.9632	-11.3475	-2.5686	1.000
	BP	-16.6712	-14.2705	-12.9452	-11.3242	-2.5804	1.000
	HGr1	-7.0235	-5.5701	-4.6782	-3.4896	3.9196	0.980
	HGr2	-3.7598	-2.9445	-2.3261	-1.3899	5.4418	0.915
	HGr3	-6.2534	-4.8622	-3.9865	-2.8122	4.6209	0.962
0.05	ZS	-12.4218	46.5266	84.9298	138.3743	472.8710	0.982
	BP	-12.4045	46.5412	84.9377	138.3800	472.8728	0.982
	HGr1	-4.3019	50.7902	88.4743	141.2488	474.0180	0.997
	HGr2	-2.0405	51.5953	89.0583	141.6796	474.0419	0.999
	HGr3	-3.5681	51.4259	89.0556	141.8120	474.4255	0.998
0.20	ZS	-4.8304	211.8973	337.9223	499.187	707.5502	0.999
	BP	-4.8135	211.8972	337.9383	499.199	709.6526	0.999
	HGr1	1.9462	214.1657	339.5490	500.279	708.1978	1.000
	HGr2	3.5341	214.4562	339.6827	500.289	708.1055	1.000
	HGr3	2.6515	214.6885	340.0109	500.670	708.5138	1.000

Table 3: Five number summary and the RF of choosing the correct model under  $\log(\text{BF})$  associated with each prior when  $\{a = 2, b = 3, m = 400\}$ .

$g$	prior	Min	$Q_1$	Median	$Q_3$	Max	RF(%)
0	ZS	-Inf	-Inf	-Inf	-Inf	-Inf	1.000
	BP	-Inf	-Inf	-Inf	-Inf	-720.8974	1.000
	HGr1	-6.1992	-4.3584	-3.3993	-2.1222	4.0161	0.937
	HGr2	-3.4643	-2.2161	-1.5290	-0.5166	5.2688	0.820
	HGr3	-5.7210	-3.6464	-2.7296	-1.4632	4.6139	0.895
0.05	ZS	-Inf	-Inf	-Inf	-Inf	-Inf	0.000
	BP	-611.9589	-499.1527	-466.2511	-437.5138	-297.2645	0.000
	HGr1	34.0548	80.6616	96.5540	111.0798	190.0442	1.000
	HGr2	34.8696	81.2014	97.0909	111.6091	190.4298	1.000
	HGr3	34.6502	81.2911	97.1527	111.6971	190.6146	1.000
0.20	ZS	-Inf	-Inf	-Inf	-Inf	-355.8564	0.000
	BP	-58.6632	179.7255	259.7000	337.8486	609.0160	0.993
	HGr1	346.8938	522.8598	585.0017	646.9002	Inf	1.000
	HGr2	347.1775	522.9843	584.9892	646.8955	Inf	1.000
	HGr3	347.6076	523.3627	585.4013	647.3217	Inf	1.000

Table 4: Five number summary and the RF of choosing the correct model under  $\log(\text{BF})$  associated with each prior when  $\{a = 30, b = 30, m = 6\}$ . The symbol 'Inf' means that the Bayes factor is less than the machine epsilon in R.

$\text{BF}^{\text{ZS}}$  in terms of the RF of the correct model. These findings exactly match the claims in Theorems 5 and 6. Even when both  $k$  and  $r$  are large (not shown here),  $\text{BF}^{\text{ZS}}$  and  $\text{BF}^{\text{BP}}$  could still be negative occasionally and support the null model for  $g = 0.05$  (i.e., the sampling model is  $M_\gamma$ ). This undesirable situation occurs, mainly because the two Bayes factors put much evidence in supporting the null model.

Numerical results in the two tables also show that  $\text{BF}^{\text{ZS}}$  and  $\text{BF}^{\text{BP}}$  have a similar behavior and both are strongly biased toward the null model ( $g = 0$ ), leading to a smaller Type I error than the proposed Bayes factors. Consequently, they perform more poorly than the proposed Bayes factors when sampling from the alternative model. This is the price that they have to pay for their smaller Type I error behavior. We observe that the proposed Bayes factors display a more balanced Type I and II error probabilities than the ones based on the ZS prior and the BP prior. We have also conducted simulation studies with other choices of  $(a, b, m)$  for the hypothesis testing problem in other ANOVA settings. Numerical findings are in good agreement with the findings mentioned above and are thus not shown here for simplicity.

## 6 Concluding remarks

We have studied Bayes factor consistency under the various mixtures of  $g$ -priors for the hypothesis testing problem in the multi-way ANOVA models with a diverging number of parameters. It has been shown that the Bayes factors based on the ZS prior and the BP prior are not always consistent due to the presence of an inconsistency region around the null model when  $k$  grows proportionally to  $n$ . The Bayes factor based on the proposed family of hyper-priors avoids this undesirable inconsistency problem. Simulation results show that the proposed Bayes factors perform well and yield satisfactory results in terms of balancing Type I and II error probabilities under different simulation situations. Among the three special choices of the proposed prior, we have a preference for the HGr1 because its overall performance is superior than the other two when sampling from the null model and they all behave similarly when sampling from the alternative model. In ongoing work, we are further studying the performance of the Bayes factors with other choices of  $f_r(t)$  given by (24).

The robust prior (Bayarri et al., 2012) has recently received much attention in context of normal linear regression models, because it was originally developed based on a number of theoretical arguments. It includes the hyper- $g$  prior and the hyper- $g/n$  prior (Liang et al., 2008) as particular cases and is also closely connected with the ZS prior studied in this paper. These observations motivate us to study Bayes factor consistency under the robust prior in the ANOVA models with a growing number of parameters, which is currently under investigation and will be reported elsewhere.

## Supplementary Material

Supplementary material for “Mixtures of  $g$ -priors for analysis of variance models with a diverging number of parameters” (DOI: [10.1214/16-BA1011SUPP](https://doi.org/10.1214/16-BA1011SUPP); .pdf).

## References

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). “Criteria for Bayesian model choice with application to variable selection.” *Annals of Statis-*

- tics*, 40(3): 1550–1577. MR3015035. doi: <http://dx.doi.org/10.1214/12-AOS1013>. 514, 518, 529
- Bayarri, M. J. and García-Donato, G. (2007). “Extending conventional priors for testing general hypotheses in linear models.” *Biometrika*, 94(1): 135–152. MR2367828. doi: <http://dx.doi.org/10.1093/biomet/asm014>. 514, 517, 518, 519
- Becker, M. and Klößner, S. (2013). *PearsonDS: Pearson Distribution System*. R package version 0.97. URL <http://CRAN.R-project.org/package=PearsonDS> 521
- Berger, J. O., Ghosh, J. K., and Mukhopadhyay, N. (2003). “Approximations and consistency of Bayes factors as model dimension grows.” *Journal of Statistical Planning and Inference*, 112(1-2): 241–258. MR1961733. doi: [http://dx.doi.org/10.1016/S0378-3758\(02\)00336-1](http://dx.doi.org/10.1016/S0378-3758(02)00336-1). 524, 525
- Berger, J. O. and Pericchi, L. R. (1996). “The intrinsic Bayes factor for model selection and prediction.” *Journal of the American Statistical Association*, 91(433): 109–122. MR1394065. doi: <http://dx.doi.org/10.2307/2291387>. 518
- Berger, J. O. and Pericchi, L. R. (2001). “Objective Bayesian methods for model selection: introduction and comparison.” In *Model selection*, volume 38 of *IMS Lecture Notes Monogr. Ser.*, 135–207. Beachwood, OH: Inst. Math. Statist. With discussion by J. K. Ghosh, Tapas Samanta, and Fulvio De Santis, and a rejoinder by the authors. MR2000753. doi: <http://dx.doi.org/10.1214/lnms/1215540968>. 512
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). “The horseshoe estimator for sparse signals.” *Biometrika*, 97(2): 465–480. MR2650751. doi: <http://dx.doi.org/10.1093/biomet/asq017>. 521
- Draper, N. and Smith, H. (1998). *Applied Regression Analysis*. New York: Wiley-Interscience. MR1614335. doi: <http://dx.doi.org/10.1002/9781118625590>. 512
- Fernández, C., Ley, E., and Steel, M. F. J. (2001). “Benchmark priors for Bayesian model averaging.” *Journal of Econometrics*, 100(2): 381–427. MR1820410. doi: [http://dx.doi.org/10.1016/S0304-4076\(00\)00076-2](http://dx.doi.org/10.1016/S0304-4076(00)00076-2). 523
- Fujikoshi, Y. (1993). “Two-way ANOVA models with unbalanced data.” *Discrete Mathematics*, 116(1-3): 315–334. MR1222130. doi: [http://dx.doi.org/10.1016/0012-365X\(93\)90410-U](http://dx.doi.org/10.1016/0012-365X(93)90410-U). 512, 516
- George, E. I. and Foster, D. P. (2000). “Calibration and empirical Bayes variable selection.” *Biometrika*, 87(4): 731–747. MR1813972. doi: <http://dx.doi.org/10.1093/biomet/87.4.731>. 518
- Girón, F. J., Moreno, E., Casella, G., and Martínez, M. L. (2010). “Consistency of objective Bayes factors for nonnested linear models and increasing model dimension.” *Revista de la Real Academia de Ciencias Exactas, Físicas Y Naturales. Serie A, Matemáticas (Ed. Impresa)*, 104(1): 57–67. MR2666441. doi: <http://dx.doi.org/10.5052/RACSAM.2010.06>. 514, 525
- Guo, R. and Speckman, P. L. (2009). “Bayes factor consistency in linear models.” In *the 2009 International Workshop on Objective Bayes Methodology, Philadelphia, June 5-9, 2009*. 512

- Jeffreys, H. (1961). *Theory of Probability*. Statistics and Computing, 3rd edn. London: Oxford University Press. MR0187257. 513, 519, 523
- Johnson, V. E. and Rossell, D. (2012). “Bayesian model selection in high-dimensional settings.” *Journal of the American Statistical Association*, 107(498): 649–660. MR2980074. doi: <http://dx.doi.org/10.1080/01621459.2012.682536>. 514
- Kass, R. E. and Raftery, A. E. (1995). “Bayes factors.” *Journal of the American Statistical Association*, 90(430): 773–795. MR3363402. doi: <http://dx.doi.org/10.1080/01621459.1995.10476572>. 512, 513
- Lazic, S. (2008). “Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology.” *BMC Physiology*, 8(1): 1–7. doi: <http://dx.doi.org/10.1186/1472-6793-8-16>. 511
- Ley, E. and Steel, M. F. J. (2012). “Mixtures of  $g$ -priors for Bayesian model averaging with economic applications.” *Journal of Econometrics*, 171(2): 251–266. MR2991863. doi: <http://dx.doi.org/10.1016/j.jeconom.2012.06.009>. 513, 519, 520
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). “Mixtures of  $g$  priors for Bayesian variable selection.” *Journal of the American Statistical Association*, 103(481): 410–423. MR2420243. doi: <http://dx.doi.org/10.1198/016214507000001337>. 513, 519, 521, 523, 526, 527, 529
- Maruyama, Y. (2012). “A Bayes factor with reasonable model selection consistency for ANOVA model.” *ArXiv:0906.4329v2 [stat.ME]*. 512, 514, 519
- Maruyama, Y. and George, E. I. (2011). “Fully Bayes factors with a generalized  $g$ -prior.” *Annals of Statistics*, 39(5): 2740–2765. MR2906885. doi: <http://dx.doi.org/10.1214/11-AOS917>. 513, 514, 519, 520
- Min, X. and Sun, D. (2015). “Bayesian model selection for a linear model with grouped covariates.” *Annals of the Institute of Statistical Mathematics*, 1–27. doi: <http://dx.doi.org/10.1007/s10463-015-0518-9>. 527
- Moreno, E., Girón, J., and Casella, G. (2015). “Posterior model consistency in variable selection as the model dimension grows.” *Statistical Science*, 30(2): 228–241. MR3353105. doi: <http://dx.doi.org/10.1214/14-STS508>. 520
- Moreno, E., Girón, F. J., and Casella, G. (2010). “Consistency of objective Bayes factors as the model dimension grows.” *Annals of Statistics*, 38(4): 1937–1952. MR2676879. doi: <http://dx.doi.org/10.1214/09-AOS754>. 514
- Pearson, K. (1895). “Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material.” 186: 343–414. 521
- Qian, S. and Shen, Z. (2007). “Ecological applications of multilevel analysis of variance.” *Ecology*, 88(10): 2489–2495. 511
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org/> 512

- Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). “Default Bayes factors for ANOVA designs.” *Journal of Mathematical Psychology*, 56(5): 356–374. MR2983394. doi: <http://dx.doi.org/10.1016/j.jmp.2012.08.001>. 511, 512, 514, 519
- Searle, S., Casella, G., and McCulloch, C. (1992). *Variance Components*. Wiley, New York. MR1190470. doi: <http://dx.doi.org/10.1002/9780470316856>. 516
- VanLeeuwen, D. (1997). “Assessing reliability of measurements with generalizability theory: an application to inter-rater reliability.” *Journal of Agricultural Education*, 38(3): 36–42. 511
- Wang, M. (2016). “Supplementary material for “Mixtures of  $g$ -priors for analysis of variance models with a diverging number of parameters”” *Bayesian Analysis*. doi: <http://dx.doi.org/10.1214/16-BA1011SUPP>. 515
- Wang, M. and Sun, X. (2013). “Bayes factor consistency for unbalanced ANOVA models.” *Statistics*, 47(5): 1104–1115. MR3175737. doi: <http://dx.doi.org/10.1080/02331888.2012.694445>. 512, 514, 523
- Wang, M. and Sun, X. (2014). “Bayes factor consistency for nested linear models with a growing number of parameters.” *Journal of Statistical Planning and Inference*, 147: 95–105. MR3151848. doi: <http://dx.doi.org/10.1016/j.jspi.2013.11.001>. 514
- Wetzels, R., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2012). “A default Bayesian hypothesis test for ANOVA designs.” *The American Statistician*, 66(2): 104–111. MR2968006. doi: <http://dx.doi.org/10.1080/00031305.2012.695956>. 512, 514, 519
- Womack, A. J., León-Novelo, L., and Casella, G. (2014). “Inference from intrinsic Bayes’ procedures under model selection and uncertainty.” *Journal of the American Statistical Association*, 109(507): 1040–1053. MR3265679. doi: <http://dx.doi.org/10.1080/01621459.2014.880348>. 525
- Zellner, A. (1986). “On assessing prior distributions and Bayesian regression analysis with  $g$ -prior distributions.” In *Bayesian inference and decision techniques*, volume 6 of *Stud. Bayesian Econometrics Statist.*, 233–243. North-Holland, Amsterdam. MR0881437. 519
- Zellner, A. and Siow, A. (1980). “Posterior odds ratios for selected regression hypotheses.” *Trabajos de Estadística Y de Investigación Operativa*, 31(1): 585–603. doi: <http://dx.doi.org/10.1007/BF02888369>. 513, 514

### Acknowledgments

The authors would like to thank the Editor, an Associate editor, and one referee for their thorough and constructive comments, which has substantially improved the quality of the paper.