Article
Informative *g*-Priors for Mixed Models

Yu-Fang Chien [1], Haiming Zhou [1,*,†], Timothy Hanson [2] and Theodore Lystig [3]

[1] Department of Statistics and Actuarial Science, Northern Illinois University, DeKalb, IL 60115, USA
[2] Structural Heart & Aortic, Medtronic, Minneapolis, MN 55432, USA
[3] BridgeBio, Palo Alto, CA 94304, USA
* Correspondence: haiming2019@gmail.com
† Current Affiliation: Daiichi Sankyo, Inc., Basking Ridge, NJ 07920, USA.

Abstract: Zellner's objective *g*-prior has been widely used in linear regression models due to its simple interpretation and computational tractability in evaluating marginal likelihoods. However, the *g*-prior further allows portioning the prior variability explained by the linear predictor versus that of pure noise. In this paper, we propose a novel yet remarkably simple *g*-prior specification when a subject matter expert has information on the marginal distribution of the response y_i. The approach is extended for use in mixed models with some surprising but intuitive results. Simulation studies are conducted to compare the model fitting under the proposed *g*-prior with that under other existing priors.

Keywords: prior elicitation; *g*-priors; linear regression; Bayesian model selection; mixed models; variable selection; Bayes factor

1. Introduction

Incorporation of expert opinion has been an integral component of informative priors for Bayesian models in a wide variety of settings, many of them clinical [1–3]. Even in a highly regulated industry such as the medical devices field, guidance has existed for some time for how expert opinion might be incorporated into models [4]. However, the willingness of regulators to accept expert opinion does not necessarily mean that the process of obtaining and utilizing such information is straightforward. Existing approaches for leveraging prior opinions tend to be cumbersome and labor intensive [5–8]. This paper provides a simple and easy to use method for experts to specify *g*-priors for a wide class of mixed models focusing only on the marginal distribution of population responses $y_1, \ldots, y_n$.

A linear model is initially considered $y_i = \mathbf{x}_i'\beta + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} (0, \sigma^2)$, and prior information (m, v) is included such that, marginally, $E(y_i) = m$ and $\mathrm{var}(y_i) = v$. Here, the notation of $x \sim (\mu, \tau^2)$ denotes that a random variable x has mean μ and variance τ^2, y_i is the *i*th response, $\mathbf{x}_i$ is a *p*-vector of covariates which usually includes an intercept, and $\beta = (\beta_1, \ldots, \beta_p)'$ is the *p*-vector of regression coefficients. The errors ϵ_i are assumed Gaussian for the bulk of the paper, but this assumption can often be relaxed. Zellner's *g*-prior [9,10] posits

$$\beta \sim N_p(\beta_0, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1}),$$

where $\mathbf{X}$ is the usual $n \times p$ design matrix, yielding a posterior mean that is a weighted average of the usual ordinary least squares (OLS) estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and the prior value β_0, i.e., $\tilde{\beta} = \frac{g}{1+g}\hat{\beta} + \frac{1}{1+g}\beta_0$. Note that $g = 0$ gives no weight to the outcome data $\mathbf{y} = (y_1, \ldots, y_n)'$ and $g \to \infty$ gives complete weight to the data. The choice of g has received considerable interest in the literature, and the *g*-prior has been widely adopted for use in variable selection e.g., [11,12]. It is not our intent to add to the burgeoning literature on variable selection here but rather provide a useful prior for model parameters when some

information about the data generating mechanism is known; in such cases, the "informative *g*-prior" developed here is competitive with existing approaches for variable selection (Section 4.1.2). In this paper, we propose an informative *g*-prior that can be used by default when prior information is lacking, or that can reflect available prior information on the marginal distribution of population responses. For example, if the outcome is cholesterol level in a certain population and the interest is to investigate how cholesterol level changes with covariates such as age, gender, ethnicity and body mass index, the expert might find that, marginally, $y_i \sim N(m = 190, v = 5^2)$ from various studies. This marginal prior specification does not rely on any covariates, which makes the prior excitation relatively easy. The theoretical marginal distribution of the y_i's can be obtained from the population distribution of covariates $\mathbf{x}_i$, the distribution on β, and the value of σ^2 through the linear regression model under a specific form of the *g*-prior. Then, the *g*-prior can be derived by matching moments between this theoretical marginal distribution and the prior distribution of $N(190, 5^2)$ to ensure, e.g., $E(y_i) = 190$ and $\text{var}(y_i) = 5^2$. The method is further extended to provide default priors for mixed models, allowing for random-effects ANOVA, random coefficient models, etc.

The sampling distribution of the OLS estimator $\hat{\beta}$ has covariance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. In a Bayesian analysis assuming normal errors, the flat prior $p(\beta) \propto 1$ yields the conditional posterior $\beta | \mathbf{y}, \mathbf{X}, \sigma^2 \sim N_p(\hat{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$. In either case, the covariance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ estimates $\frac{\sigma^2}{n}[\mu\mu' + \Sigma]^{-1}$, where $E(\mathbf{x}_i) = \mu$ and $\text{cov}(\mathbf{x}_i) = \Sigma$. That is, greater variability in $\mathbf{x}_i$ implies greater precision in estimating β. Thus, ref. [9] specifies a vague conditional prior for β that takes advantage of information on distributional shape based solely on $\mathbf{X}$ and a flat prior on σ^2. The *g*-prior developed here further separates how much marginal variability in y_i is explained *a priori* by the model from that of pure noise σ^2; a default specification assumes a flat uniform prior on this quantity.

Two popular classes of priors for regression models are conditional means priors [13,14] and power priors [15]. Conditional means priors require a subject matter expert to provide information on the mean response for several candidate vectors of covariates (that do not have to be among those actually observed); the usual specification requires the expert to be able to think about the mean responses independently, but this is not strictly required. Let the candidate vectors be $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N$, where $N \leq p$. The subject matter expert is asked to provide, say, a 95% interval that contains the mean response $\tilde{\mathbf{x}}_i'\beta$ at covariates $\tilde{\mathbf{x}}_i$, e.g., $P(a_i \leq \tilde{\mathbf{x}}_i'\beta \leq b_i) = 0.95$. This information on the conditional means $\tilde{m}_i = \tilde{\mathbf{x}}_i'\beta$ is summarized as $\tilde{m}_i \overset{ind.}{\sim} N(m_i, v_i)$, yielding $\tilde{\mathbf{m}} = \tilde{\mathbf{X}}\beta \sim N_N(\mathbf{m}, \mathbf{V})$, where $\mathbf{m} = (m_1, \ldots, m_N)$ and $\mathbf{V} = \text{diag}(v_1, \ldots, v_N)$. If $\tilde{\mathbf{X}}$ is invertible, requiring $N = p$; then, the induced prior is simply $\beta \sim N_N(\tilde{\mathbf{X}}^{-1}\mathbf{m}, \tilde{\mathbf{X}}^{-1}\mathbf{V}(\tilde{\mathbf{X}}^{-1})')$. Ref. [13] propose methods for handling partial prior information on a subset $N < p$, i.e., the subject matter expert only need specify a handful of priors for conditional means. In contrast, the *g*-prior developed here only requires information on the marginal distribution of the y_i's, namely (m, v).

Power priors are built from historical regression data having the same covariates as the current data. Say that the historical data are $\{\tilde{\mathbf{x}}_i, \tilde{y}_i\}_{i=1}^M$ and the current data are $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$. The power prior is simply the posterior of β based on a reference prior, raised to the power $\alpha \in [0, 1]$: $p(\beta, \sigma^2) \propto \left[\prod_{i=1}^M \phi(\tilde{y}_i | \tilde{\mathbf{x}}_i'\beta, \sigma^2)\right]^\alpha \sigma^{-2}$, where $\phi(y|m, v)$ is the density of a normal random variable with mean m and variance v. The parameter α provides the "degree of borrowing" from the historical data, with $\alpha = 0$ giving none and $\alpha = 1$ treating the historical data the same as the current study data. The choice of α has also received considerable research [16–18]. In addition to the power and conditional mean priors, ref. [19] proposed a natural conjugate reference informative prior by taking into account various degrees of certainty in covariates, and [20] proposed a default prior for β_js by using a normal distribution with mean zero and standard deviation equal to the standard error of the M-estimator of each β_j.

There are several notable limitations for conditional means and power priors. Conditional means priors involve the analyst thinking about various covariate combinations

and providing information on the mean response for each covariate setting. As the number of predictors increases, this becomes increasingly difficult; it can be conceptually easier to think about marginal quantities such as the overall mean m and variance v in the population. Such marginal information may be available via census or through published summary data. The power prior requires a historical data set having a superset of the variables under consideration in the current study, which is often unavailable for new treatments.

One consequence of the proposed priors developed here is that proper, data-driven priors are given in closed-form with default settings. Thus, standard model comparison via Bayes factors is possible as no improper priors are used. Difficult-to-elicit prior information such as the range of a variance component is replaced with the question "How much variability in the data do you think the model explains?" If the answer to this is "I have no idea" then a uniform distribution on σ^2 is suggested. The proposed priors do not have closed-form full conditional distributions for all parameters but are easily specified and fit in R using statistical software Just Another Gibbs Sampler (JAGS) [21] via packages such as `R2jags` [22].

Bayesians have long known that injecting a small amount of prior information can often "fix" pathological MCMC schemes. The *g*-prior developed here can be viewed as a ridge prior that takes multicollinearity into account, with the added benefit that the ridge parameter is automatically chosen by the data. Section 2 introduces the informative *g*-prior for linear regression models. Section 3 extends the *g*-prior for use in mixed models. Section 4 presents a detailed set of simulation studies exploring the use of the *g*-prior and compares to other priors in common use. Section 5 concludes the paper with a discussion and eye toward future research.

2. Prior for Linear Regression Models

2.1. The Prior in [23]

The *g*-prior in [23] was developed for logistic regression; this section carefully extends their approach to normal-errors regression, and Section 3 generalizes further to mixed models. Their *g*-prior is specified as

$$\beta | \beta_0, g, \mathbf{X} \sim N_p(\beta_0, gn(\mathbf{X}'\mathbf{X})^{-1}), \tag{1}$$

where $g > 0$, and $\mathbf{X}$ is the usual $n \times p$ design matrix. Assume $\mathbf{x}_i \overset{iid}{\sim} H$ for some distribution H where $\mathbf{x}_i \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Noting that $\mathbf{x}_i$ includes the intercept in the first element, the first element of $\boldsymbol{\mu}$ is one and the first row along with the first column entries of $\boldsymbol{\Sigma}$ are all zeros. Given the data $\mathbf{X}$, for any new subject with response y and covariates $\mathbf{x} \sim H$, assuming $\mathbf{x}$ and β are mutually independent, one has $E(\mathbf{x}'\beta) = E_{\mathbf{x}}\{E_\beta(\mathbf{x}'\beta|\mathbf{x})\} = \boldsymbol{\mu}'\beta_0$, by the law of iterated expectations. In addition, by the law of total variance, one has

$$
\begin{aligned}
\text{Var}(\mathbf{x}'\beta) &= E_{\mathbf{x}}\{\text{Var}_\beta(\mathbf{x}'\beta|\mathbf{x})\} + \text{Var}_{\mathbf{x}}\{E_\beta(\mathbf{x}'\beta|\mathbf{x})\} \\
&= E_{\mathbf{x}}\{gn\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}\} + \text{Var}_{\mathbf{x}}(\boldsymbol{\mu}'\beta_0) \\
&= g \cdot \text{trace}\left\{ n(\mathbf{X}'\mathbf{X})^{-1}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') \right\} \\
&\overset{p}{\to} g \cdot \text{trace}\left\{ (\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')^{-1}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}') \right\} = gp,
\end{aligned}
$$

where $\overset{p}{\to}$ denotes convergence in probability, and the limiting statement originates from the fact that $n(\mathbf{X}'\mathbf{X})^{-1} \overset{p}{\to} (\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma})^{-1}$. Hence, given $\mathbf{X}$, the *g*-prior in (1) implies that $\mathbf{x}'\beta$ has a variance approximately equal to gp for any covariate $\mathbf{x}$ randomly drawn from its population H. [23] found that $\mathbf{x}'\beta$ also often approximately follows a normal distribution, and this approximation is good for a variety of H considered in their simulations, even when some covariates are categorical. Therefore, it is reasonable to assume that $\mathbf{x}'\beta$ approximately follows $N(\boldsymbol{\mu}'\beta_0, gp)$.

For the linear normal regression model $y_i|\mathbf{x}_i \overset{ind.}{\sim} N(\mathbf{x}_i'\beta, \sigma^2)$, the *g*-prior in [23] can be applied as follows. Assuming a subject matter expert has in hand information on the distribution of marginal mean responses (i.e., $E(y_i) \overset{set}{=} m$) in a population, rather than the distribution of y_i, say $m \sim N(\mu_m, \sigma_m^2)$ with (μ_m, σ_m^2) being chosen to reflect the prior knowledge about the distribution of m. Then using the prior matching idea in [23], one can immediately solve for β_0 and g in (1) as $\beta_0 = \mu_m \mathbf{e}_1$ and $g = \sigma_m^2/p$ where $\mathbf{e}_1 = (1, 0, \dots, 0)'$. Although [24] finds the default prior given by [23] for logistic regression to provide the best predictive performance among several contenders, the performance for linear regression model has not been well tested. In addition, it is not straightforward to set default values for (μ_m, σ_m^2) and its extension to linear mixed models is not readily available.

In this paper, we will propose a new *g*-prior for the linear regression model when a subject matter expert has information on the marginal distribution of the response y_i rather than $E(y_i)$ with reasonable default settings and then extend it for use in linear mixed models.

2.2. New Prior Development

An easily implemented *g*-prior is first proposed for use in the linear regression model:

$$y_i|\mathbf{x}_i, \beta, \sigma^2 \overset{ind.}{\sim} (\mathbf{x}_i'\beta, \sigma^2), \ i = 1, \dots, n. \tag{2}$$

Consider the situation where a subject matter expert has information on the *marginal* distribution of observations y_i that can be synthesized as

$$y_i \sim (m, v), \ m \sim N(m_0, v/k_m), \ v^{-1} \sim \Gamma(k_v, v_0 k_v), \tag{3}$$

where m_0, v_0, k_m and k_v can be obtained from previous studies or published summary data; details are given in Section 2.3. Here $\Gamma(a, b)$ denotes the gamma distribution with mean equal to a/b. The goal here is to develop a particular version of *g*-prior on (β, σ^2) in (2) that achieves the marginal distribution of $y_i \sim (m, v)$.

Consider the *g*-prior in (1). Given σ^2, the total expectation formula gives

$$E(y) = E_{\mathbf{x}'\beta} E_{y|\mathbf{x}'\beta}(y) = E_{\mathbf{x}'\beta}(\mathbf{x}'\beta) = \boldsymbol{\mu}'\beta_0,$$

and the total variance formula gives

$$\text{var}(y) = E_{\mathbf{x}'\beta} \text{var}_{y|\mathbf{x}'\beta}(y) + \text{var}_{\mathbf{x}'\beta} E_{y|\mathbf{x}'\beta}(y) = E_{\mathbf{x}'\beta}(\sigma^2) + \text{var}_{\mathbf{x}'\beta}(\mathbf{x}'\beta) \approx \sigma^2 + gp.$$

For models with an intercept, setting $\beta_0 = m\mathbf{e}_1$ satisfies the first moment condition $E(y_i) = m$. The larger σ^2 is, the more the prior shrinks β toward the intercept only model (with an intercept focused on m), and so is conservative in favoring the null of the overall *F*-test that no covariates are important.

To match the second moment condition $\text{var}(y_i) = v$, set $gp + \sigma^2 = v$ and solve for $g = (v - \sigma^2)/p$ in (1) when $\sigma^2 \leq v$. Since $E(y_i|\sigma, m, v) = m$ and $\text{var}(y_i|\sigma, m, v) = v$ for all $\sigma \geq 0$, the marginal constraint of $y_i \sim (m, v)$ approximately holds for any prior $\sigma^2 \sim p(\cdot)$ with support $\sigma^2 \in [0, v]$. In particular, a special case of the generalized beta distribution,

$$p_{a,b,v}(\sigma^2) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)v} \left(\frac{\sigma^2}{v}\right)^{a-1} \left(1 - \frac{\sigma^2}{v}\right)^{b-1} I_{[0,v]}(\sigma^2), \tag{4}$$

denoted $\text{gb}(a, b, v)$, allows flexibility in specifying how much variability the regression model explains relative to the total variability v; note $E[\frac{v-\sigma^2}{v}|v] = \frac{b}{a+b}$. If one had prior information that, say, the amount of variation explained by regression is r_0 (similar to R^2 in OLS regression, but R^2 conditions on $\mathbf{X}$ and fixes $\beta = \hat{\beta}$), then the parameters in (4) could be chosen such that $\frac{b}{a+b} = r_0$ with the total "sample size" n_0 going into the prior as

$n_0 = a + b$; solving yields $a = (1 - r_0)n_0$ and $b = r_0 n_0$. No prior preference gives $a = b = 1$, i.e., $\sigma^2 \sim \text{uniform}(0, v)$, a sensible default choice.

Encapsulating the above, a hierarchical prior that maintains $y_i \sim (m, v)$ is

$$\beta | \sigma^2, m, v \sim N_p\left(\mathbf{e}_1 m, \tfrac{n}{p}(v - \sigma^2)(\mathbf{X}'\mathbf{X})^{-1}\right),\ \sigma^2 | v \sim \text{gb}(a, b, v),$$
$$m \sim N(m_0, v/k_m),\ v^{-1} \sim \Gamma(k_v, v_0 k_v). \tag{5}$$

This prior provides an intuitive interpretation given v: when $\sigma^2 = 0$ ($b \to \infty$) the model explains all variability in y_i, when $\sigma^2 = v$ ($a \to \infty$) then the model explains nothing. Values $a, b \in (0, \infty)$ indicates that the truth is somewhere between these two extremes, with $a = b = 1$ reflecting no preference on how much variability the model explains. This formulation of the *g*-prior can be viewed as a type of ridge regression which further addresses multicollinearity among predictors, but where the ridge parameter is chosen automatically. The special form of the *g*-prior enables easy computation of the amount of variability the model explains relative to the total v.

Once a distribution (e.g., Gaussian) is assumed for the linear regression model given in (2), estimates of β and σ^2 can be obtained using statistical software such as JAGS [21] via the R package `R2jags` [22]; see Supplementary Materials for the R code.

2.3. Hyper-Prior Elicitation for (M, V)

Our prior in (5) requires a specification for hyperparameters m_0, v_0, k_m and k_v. Suppose we have historical data $\mathbf{y}_o = (y_{o1}, \dots, y_{oM})$ from a similar study population. If we assume $y_{oi} \overset{iid}{\sim} N(m, v)$, then using a noninformative prior such as $p(m, v) \propto 1/\sqrt{v}$ gives

$$m | \mathbf{y}_o, v \sim N(\bar{y}_o, v/M),\ v^{-1} | \mathbf{y}_o \sim \Gamma(M/2, s_{y_o}^2 \cdot M/2), \tag{6}$$

where $s_{y_o}^2 = \sum_{i=1}^n (y_{oi} - \bar{y}_o)^2 / M$. If one believes that the historical data $\mathbf{y}_o$ come from the same population as the current observed response data $\mathbf{y} = (y_1, \dots, y_n)$, it is reasonable to set $m_0 = \bar{y}_o$, $v_0 = s_{y_o}^2$, $k_m = M$ and $k_v = M/2$ in (5). If the historical data come from a population quite different from the current study or the population distribution is not plausibly normal, one may set lower values for k_m and k_v to put less weight on the historical data relative to the current data. If historical data are not available, we recommend setting $m_0 = \bar{y} = \sum_{i=1}^n y_i/n$, $v_0 = s_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2/n$, $k_m = 2$ and $k_v = 1$ instead of setting $k_m = k_v = 0$; this assumes that the unavailable historical data have the sample mean equal to $\bar{y}$ and the sample variance equal to s_y^2 and are given with the weight of two observations. In real applications, a sensitivity analysis can be performed by setting k_m to several different values between 2 and M.

The idea behind our hyperprior elicitation for (m, v) is similar to the power prior [15], which is defined as the posterior of model parameters given the historical data, raised to a power $\alpha \in [0, 1]$, where α provides the "degree of borrowing" from the historical data. Consider an intercept-only model $y_i \overset{iid}{\sim} N(\beta_1, \sigma^2)$. Note that our prior (5) simply reduces to $\beta_1 = m$, $\sigma^2 = v$, $m \sim N(m_0, v/k_m)$, $v^{-1} \sim \Gamma(k_v, v_0 k_v)$. Given the historical data $\mathbf{y}_o$, setting $m_0 = \bar{y}_o$, $v_0 = s_{y_o}^2$, $k_m = M$ and $k_v = M/2$ is exactly the power prior with $\alpha = 1$. Similarly, the values of k_m and k_v control the influence of historical data. For the general linear model in (2), the important difference is that our prior on (β, σ^2) does not require any covariates in the historical data, since it depends on the historical data only through (m, v).

2.4. Comparing to the Mixture of G Priors

For the linear model in (2) with Gaussian errors, the hyper-g prior in [12] can be expressed as

$$\beta|g,\sigma^2 \sim N_p(m\mathbf{e}_1, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$
$$p(m,\sigma^2) \propto 1/\sigma^2$$
$$p(g) = \frac{h-2}{2}(1+g)^{-h/2}I(g>0),$$

(7)

where $h > 2$ is set to ensure a proper distribution. Ref. [12] show that the hyper g-prior is not consistent for model selection when the true model is the null model, and then propose a hyper-g/n prior as

$$p(g) = \frac{h-2}{2n}(1+g/n)^{-h/2}I(g>0).$$

(8)

Setting $g\sigma^2 = \frac{n}{p}(v-\sigma^2)$ i.e., $\sigma^2/v = \frac{1}{gp/n+1}$ in our prior (5) gives

$$\beta|g,\sigma^2 \sim N_p\left(\mathbf{e}_1 m, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right),$$
$$m|g,\sigma^2 \sim N(m_0, \sigma^2(gp/n+1)/k_m),$$
$$\sigma^{-2}|g \sim \Gamma(k_v, k_v \frac{v_0}{gp/n+1}),$$
$$\frac{1}{gp/n+1} \sim \text{beta}(a,b).$$

(9)

If we set $a = \frac{h-2}{2}$, $b = 1$, $k_m = 0$ and $k_v = 0$, it is easy to show that our prior further becomes

$$\beta|g,\sigma^2 \sim N_p\left(\mathbf{e}_1 m, g\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right),$$
$$p(m,\sigma^2) \propto 1/\sigma^2,$$
$$p(g) = \frac{h-2}{2n/p}(1+gp/n)^{-h/2}I(g>0).$$

(10)

which is similar to the hyper-g/n prior in (8), the only difference being that our g is scaled by n/p instead of n. Therefore, the proposed prior here naturally leads to a modified version of the hyper-g/n prior considered in [12] when there is no history information on (m, v).

2.5. Simple Example

Ref. [25] analyze data on the $n = 27$ lengths y_i (in meters) of dugongs (sea cows) having ages a_i (in years). They fit a nonlinear exponential model for length based on a_i; we consider a linear model by transforming age, i.e., $\mathbf{x}_i = (1, \log(a_i))'$. An example of a commonly used vague, proper prior is $\beta_0, \beta_1 \overset{iid}{\sim} N(0, 10^2)$ and $\sigma^{-2} \sim \Gamma(0.01, 0.01)$. The prior marginal mean and variance for the response y under this prior can be estimated via Monte Carlo (MC) by simulating $\sigma^{-2} \sim \Gamma(0.01, 0.01)$, $\beta \sim N_2(\mathbf{0}, 10^2\mathbf{I}_2)$, and $y_i^{(l)} \overset{iid}{\sim} N(\mathbf{x}_i'\beta, \sigma^2)$, $i = 1, \ldots, 27$, $l = 1, \ldots, 1000$, yielding 1000 datasets $\{y_i^{(l)} : i = 1, \ldots, n\}$. The simulation of $\sigma^{-2} \sim \Gamma(0.01, 0.01)$ is completed using the method of [26] designed for gamma distributions with small shape parameters. The average prior sample mean (across the 1000 datasets) and prior sample variance are around 2×10^{120} and 2×10^{249}, respectively. These are nowhere near the observed sample mean and variance of $\bar{y} = 2.334$ and $s_y^2 = 0.073$. In contrast, a similar simulation under our proposed new g-prior in (5) with $a = b = 1$, $m_0 = \bar{y}$, $v_0 = s_y^2$, $k_m = 2$ and $k_v = 1$ yields an average sample mean of 2.305 with MC standard deviation of 0.559 and an average sample variance of 0.442 with MC standard deviation of 3.243. That is, the inference under our prior focuses on a much smaller set of potential models that could have conceivably generated the observed data. The posterior estimates for β_0, β_1, and σ^2 under our proposed new g-prior are 1.770 (0.047), 0.273 (0.021),

and 0.0094 (0.0029), respectively, where the values in parentheses are posterior standard deviations. The commonly used vague priors specified above yield to similar estimates but with slightly higher posterior standard deviations: 1.763 (0.047), 0.277 (0.021), and 0.0097 (0.0031).

The use of such prior predictive checks have recently been advocated by [27–29]; in particular, ref. [27] suggests that analysts "...visualize simulations from the prior marginal distribution of the data to assess the consistency of the chosen priors with domain knowledge." They further suggest the use of "weakly informative" priors to gently urge the prior in the direction of providing plausible marginal values. This requires some thought and visual exploration on the part of the user; the prior developed here provides a safe, default method for nudging the prior toward domain knowledge in the form of either historical marginal values or the sample moments seen in the data. The prior mean and variance exist whether analyst wants to think about them or not; this example illustrates that "vague" priors are not necessarily noninformative.

2.6. Variable Selection

Consider the Gaussian linear regression model $\mathbf{y} \sim N_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$. Using the proposed *g*-prior in (5) for Bayesian variable selection requires the calculation of marginal likelihood for each of the 2^{p-1} submodels, denoted as $\mathcal{M}_\xi$, where $\xi = (\xi_1, \ldots, \xi_p)' \in \{0,1\}^p$ is a *p*-dimensional vector of indicators with $\xi_j = 1$ implying that the *j*th covariate x_{ij} is included in the model. Here we always set $\xi_1 = 1$ so that an intercept is included. Under model $\mathcal{M}_\xi$, we have $\mathbf{y} \sim N_n(\mathbf{X}_\xi\beta_\xi, \sigma^2\mathbf{I}_n)$, where $\mathbf{X}_\xi$ is the $n \times p_\xi$ design matrix under model $\mathcal{M}_\xi$, and β_ξ is the corresponding p_ξ-vector of regression coefficients. For model $\mathcal{M}_\xi$, a default prior specification for β_ξ and σ^2 is given by

$$\beta_\xi | \sigma^2 \sim N\left(m\mathbf{e}_{1\xi}, \frac{n}{p_\xi}(v - \sigma^2)(\mathbf{X}'_\xi\mathbf{X}_\xi)^{-1}\right), \ \sigma^2 \sim \text{gb}(a, b, v),$$
$$m \sim N(m_0, v/k_m), \ v^{-1} \sim \Gamma(k_v, v_0 k_v),$$
(11)

where $\mathbf{e}_{1\xi} = (1, 0, \ldots, 0)'$ is p_ξ-dimensional.

To perform the variable selection, we need to calculate the Bayes factors for comparing each model $\mathcal{M}_\xi$ with the null model $\mathcal{M}_N = \mathcal{M}_{\mathbf{e}_1}$. Note that under model $\mathcal{M}_\xi \ (\neq \mathcal{M}_N)$ with prior (11), the marginal likelihood given σ^2 and (m, v) is

$$p(\mathbf{y}|\mathcal{M}_\xi, \sigma^2, m, v) = (2\pi)^{-\frac{n}{2}} \left[\frac{p_\xi}{nv - (n - p_\xi)\sigma^2}\right]^{p_\xi/2} (\sigma^2)^{-\frac{n - p_\xi}{2}}$$
$$\times \exp\left\{-\frac{(SST)p_\xi\left[1 + \frac{n(v - \sigma^2)(1 - R_\xi^2)}{p_\xi\sigma^2} + \frac{n(\bar{y} - m)^2}{SST}\right]}{2[nv - (n - p_\xi)\sigma^2]}\right\},$$
(12)

where $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ and R_ξ^2 is the usual R-squared under model $\mathcal{M}_\xi$. Under the null model $\mathcal{M}_N$: $y_i \sim N(\beta_1, \sigma^2)$, the prior (11) simply reduces to $\beta_1 = m$, $\sigma^2 = v$, $m \sim N(m_0, v/k_m)$, $v^{-1} \sim \Gamma(k_v, v_0 k_v)$. Therefore, the marginal likelihood under the null model given (m, v) is

$$p(\mathbf{y}|\mathcal{M}_N, m, v) = (2\pi)^{-\frac{n}{2}} (v)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v}\sum_{i=1}^n (y_i - m)^2\right\}.$$
(13)

Note that this is a special case of (12) with $\sigma^2 = v$ and $p_\xi = 1$.

When (m, v) is fixed and known, the Bayes factor for comparing any model $\mathcal{M}_\xi$ ($\neq \mathcal{M}_N$) to the null model $\mathcal{M}_N$ is

$$\mathrm{BF}[\mathcal{M}_\xi : \mathcal{M}_N | m, v] = \frac{p(\mathbf{y} | \mathcal{M}_\xi, m, v)}{p(\mathbf{y} | \mathcal{M}_N, m, v)}, \tag{14}$$

where $p(\mathbf{y} | \mathcal{M}_\xi, m, v) = \int_0^v p_{a,b,v}(\sigma^2) p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2$. It is easy to show that the Bayes factor in (14) is finite for all $p_\xi \leq p < n$. The integrals in $p(\mathbf{y} | \mathcal{M}_\xi, m, v)$ can be numerically computed using the R function `integrate` [30].

When the hyperprior on (m, v) in (5) is used, the Bayes factor for comparing $\mathcal{M}_\xi$ to $\mathcal{M}_N$ becomes

$$\mathrm{BF}[\mathcal{M}_\xi : \mathcal{M}_N] = \frac{E_{m,v}[p(\mathbf{y} | \mathcal{M}_\xi, m, v)]}{E_{m,v}[p(\mathbf{y} | \mathcal{M}_N, m, v)]}, \tag{15}$$

where the expectation $E_{m,v}[\cdot]$ is taken under the prior for (m, v) in (5). However, the calculation of expectations $E_{m,v}[\cdot]$ in (15) is considerably more computationally demanding. Based on the competitive performance of our prior compared to other methods in simulation studies, we recommend using the Bayes factor in (14) with (m, v) fixed at $(\hat{m}, \hat{v})$, where $\hat{m}$ and $\hat{v}$ are determined as follows. If there is no history information available for (m, v), we simply use $\hat{m} = \bar{y}$ and $\hat{v} = s_y^2$ based on the current marginal data $\{y_i\}$. If there is some history information for (m, v) that can be summarized as $m \sim N(m_0, v/k_m)$, $v^{-1} \sim \Gamma(k_v, v_0 k_v)$, we set $(\hat{m}, \hat{v}) = (\tilde{m}, \tilde{v})$, where $(\tilde{m}, \tilde{v})$ is the posterior mean estimate for (m, v) based on only the marginal data $(y_1, \ldots, y_n)$; see Section 2.3 for the specification of m_0, v_0, k_m and k_v with historical data $\mathbf{y}_o$. Note that closed-form formulas for $\tilde{m}$ and $\tilde{v}$ can be derived; see [31] for the derivations. Once the model is selected, we can apply the prior (5) to fit the model under the selected model.

Information Paradox

The information paradox [32] refers to the situations when we have very strong information supporting a non-null model $\mathcal{M}_\xi$, but the Bayes factor $\mathrm{BF}[\mathcal{M}_\xi : \mathcal{M}_N | m, v]$ does not go to ∞ as the information about $\mathcal{M}_\xi$ accumulate (i.e., $R_\xi^2 \to 1$). The proposed informative g prior resolves the information paradox in the sense that $\mathrm{BF}[\mathcal{M}_\xi : \mathcal{M}_N | m, v] \to \infty$ with fixed n, $p_\xi \leq p \leq (n-2)$ and $R_\xi^2 \to 1$. Note that the denominator in (14) is finite, and by the mean value theorem for definite integrals, there exists c in $(0, v)$ such that

$$\int_0^v p_{a,b,v}(\sigma^2) p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2 = p_{a,b,v}(c) \int_0^v p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2.$$

Therefore, it suffices to show that

$$\lim_{R_\xi^2 \to 1} \int_0^v p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2 = \infty \text{ for all } p_\xi \leq p \leq (n-2).$$

Noting that $\int_0^v p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2$ is an increasing function of R_ξ^2, we have

$$\lim_{R_\xi^2 \to 1} \int_0^v p(\mathbf{y} | \mathcal{M}_\xi, \sigma^2, m, v) d\sigma^2 = \int_0^v (2\pi)^{-\frac{n}{2}} \left[\frac{p_\xi}{nv - (n - p_\xi)\sigma^2} \right]^{p_\xi/2} (\sigma^2)^{-\frac{n - p_\xi}{2}}$$

$$\times \exp \left\{ -\frac{(SST) p_\xi \left[1 + \frac{n(\bar{y} - m)^2}{SST} \right]}{2[nv - (n - p_\xi)\sigma^2]} \right\} d\sigma^2$$

$$\geq \int_0^v (2\pi)^{-\frac{n}{2}} \left[\frac{p_\xi}{nv}\right]^{p_\xi/2} (\sigma^2)^{-\frac{n-p_\xi}{2}}$$

$$\times \exp\left\{ -\frac{(SST)p_\xi\left[1 + \frac{n(\bar{y}-m)^2}{SST}\right]}{2[nv - (n-p_\xi)v]} \right\} d\sigma^2$$

$$\geq constant \int_0^v (\sigma^2)^{-\frac{n-p_\xi}{2}} d\sigma^2 = \infty$$

for all $p_\xi \leq p \leq (n-2)$.

3. Mixed Models

3.1. One-Way Random Effects ANOVA

The *g*-prior developed in Section 2 for regression models can be immediately extended to mixed models in an analogous fashion. The shrinkage induced by the *g*-prior yields familiar exchangeable prior specifications already in widespread use as special cases, as well as some new default formulations. We first examine the simplest random effects model, a one-way ANOVA, typically formulated as

$$y_{ij} = \beta + \gamma_i + \epsilon_{ij}, \; i = 1, \ldots, c, \; j = 1, \ldots, n_i, \tag{16}$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, rewritten in matrix form as

$$\mathbf{y}_i = \mathbf{1}_{n_i}\beta + \mathbf{1}_{n_i}\gamma_i + \boldsymbol{\epsilon}_i, \tag{17}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})$, $\mathbf{1}_{n_i}$ is a n_i-vector of ones, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})$. Note that without further constraints, (16) is overparameterized; shrinkage on both the "fixed" and "random" portions separately is required for identifiability.

Noting that $n_i(\mathbf{1}'_{n_i}\mathbf{1}_{n_i})^{-1} = 1$ for all i, a *g*-prior on the first portion is

$$\beta|g_1 \sim N(\beta_0, g_1).$$

Similarly, a *g*-prior on the second portion is

$$\gamma_1, \ldots, \gamma_c|g_2 \overset{iid}{\sim} N(0, g_2).$$

This prior is *the same* as assuming exchangeable random effects, e.g., $\gamma_1, \ldots, \gamma_c|\sigma_r^2 \overset{iid}{\sim} N(0, \sigma_r^2)$, where $\sigma_r^2 = g_2$; placing a prior on g_2 is the same as placing a prior on σ_r^2. The *g*-prior as a ridge prior is evident here, with model identifiability achieved by shrinking γ_i towards 0. The amount of shrinkage is controlled *a priori* via the parameter g_2. There are obvious links from the *g*-prior to ridge regression, shrinkage priors, and penalized likelihood.

The prior on σ_r has received considerable interest; suggestions include the half-Cauchy prior and uniform priors [33], as well as approximations to Jeffreys' prior, e.g., $\sigma_r^{-2} \sim \Gamma(0.001, 0.001)$, which permeated Bayesian literature in the 1990's. Ref. [34] advocate a data-driven prior that is similar in spirit to what is presented here. Ref. [35] considers a shrinkage prior for σ_r^2 induced by a uniform prior on $\sigma^2/(\sigma^2 + \sigma_r^2)$. Ref. [36] uses a *g*-prior for ANOVA with diverging number of parameters. In contrast, we will build a prior that facilitates the borrowing of history information on the overall marginal mean m and variance v of the data y_{ij}.

3.2. Linear Mixed Models

Now consider the linear mixed model

$$y_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\gamma_i + \epsilon_{ij}, \text{ or equivalently, } \mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\gamma_i + \boldsymbol{\epsilon}_i, \tag{18}$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, γ_i is a k-vector of random effects, $\mathbf{X}_i = [\mathbf{x}_{i1} \cdots \mathbf{x}_{in_i}]'$, $\mathbf{Z}_i = [\mathbf{z}_{i1} \cdots \mathbf{z}_{in_i}]'$. In this setting, $i = 1, \ldots, c$ denotes the data cluster associated with γ_i and $j = 1, \ldots, n_i$ are the number of repeated measures within cluster i; the total sample size is $n = \sum_{i=1}^{c} n_i$. The variability in model (18) is portioned to $\mathbf{X}_i \beta$, $\mathbf{Z}_i \gamma_i$ and ϵ_i. The first two components will have dependent g-priors, inducing differing amounts of shrinkage across the two regression models; the second portion is further shrunk toward zero. Set $\gamma = (\gamma_1', \ldots, \gamma_c')'$. Again, the goal here is to develop a prior on $(\beta, \gamma, \sigma^2)$ that incorporates the marginal information of $y_{ij} \sim (m, v)$, where a hyperprior on (m, v) can be extracted from historical data or expert option. The usual g-prior on β for cluster i is

$$\beta | g_1 \sim N_p\left(m\mathbf{e}_1, g_1 n_i (\mathbf{X}_i' \mathbf{X}_i)^{-1}\right).$$

Let $\boldsymbol{\mu}_{ix} = E(\mathbf{x}_{ij}|i)$ and $\boldsymbol{\Sigma}_{ix} = \text{cov}(\mathbf{z}_{ij}|i)$ denote by mean and covariance of $\mathbf{x}_{ij}$ for cluster i, then $\mathbf{X}_i' \mathbf{X}_i / n_i \overset{p}{\to} (\boldsymbol{\mu}_{ix} \boldsymbol{\mu}_{ix}' + \boldsymbol{\Sigma}_{ix})$. Similarly, let $\boldsymbol{\mu}_x = E(\mathbf{x}_{ij})$ and $\boldsymbol{\Sigma}_x = \text{cov}(\mathbf{x}_{ij})$ denote the overall mean and covariance across all clusters, set $\mathbf{X} = [\mathbf{X}_1' \cdots \mathbf{X}_c']'$, then $\mathbf{X}'\mathbf{X}/n = \frac{1}{n} \sum_{i=1}^{c} \mathbf{X}_i' \mathbf{X}_i \overset{p}{\to} [\boldsymbol{\mu}_x \boldsymbol{\mu}_x' + \boldsymbol{\Sigma}_x]$. Noting that the same coefficient β is used for all clusters, the overall g-prior for β can be set as

$$\beta | g_1 \sim N_p\left(m\mathbf{e}_1, g_1 n (\mathbf{X}'\mathbf{X})^{-1}\right). \tag{19}$$

The usual g-prior on γ_i for cluster i is

$$\gamma_i | g_2 \overset{ind.}{\sim} N_k\left(\mathbf{0}, g_2 n_i (\mathbf{Z}_i' \mathbf{Z}_i)^{-1}\right) \approx N_k\left(\mathbf{0}, g_2 (\boldsymbol{\mu}_i \boldsymbol{\mu}_i' + \boldsymbol{\Sigma}_i)^{-1}\right), \tag{20}$$

where $\boldsymbol{\mu}_i = E(\mathbf{z}_{ij}|i)$, $\boldsymbol{\Sigma}_i = \text{cov}(\mathbf{z}_{ij}|i)$. Denote by $\boldsymbol{\mu} = E(\mathbf{z}_{ij})$ and $\boldsymbol{\Sigma} = \text{cov}(\mathbf{z}_{ij})$ the overall mean and covariance of $\mathbf{z}_{ij}$ across all clusters. If the $\mathbf{z}_{ij}$s come from the same population, i.e., $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_1 = \cdots \boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}$, (20) is equivalent to

$$\gamma_1, \ldots, \gamma_c | g_2 \overset{iid}{\sim} N_k(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega} = g_2 [\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}]^{-1}$. This final expression lies at the heart of hundreds of mixed model analyses; the derivation here clarifies that this is exactly what the g-prior gives us when $\mathbf{z}_{ij} \overset{iid}{\sim} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Define $\mathbf{Z} = [\mathbf{Z}_1' \cdots \mathbf{Z}_c']'$. Noting that $\mathbf{Z}'\mathbf{Z}/n = \frac{1}{n} \sum_{i=1}^{c} \mathbf{Z}_i' \mathbf{Z}_i \overset{p}{\to} [\boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\Sigma}]$, a sensible default prior is

$$\gamma_i | g_2 \overset{iid}{\sim} N_k\left(\mathbf{0}, g_2 n (\mathbf{Z}'\mathbf{Z})^{-1}\right), \tag{21}$$

assuming $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_1 = \cdots \boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}$ is approximately correct.

Let $t_k(r, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the k-dimensional multivariate t distribution with degrees of freedom r, mean $\boldsymbol{\mu}$ for $r > 1$, and covariance $\frac{r}{r-2} \boldsymbol{\Sigma}$ for $r > 2$. Taking $r/g_2 \sim \chi_r^2$ under the default prior (21), the induced marginal prior on γ_i is a multivariate t distribution see [37], given by

$$\gamma_i \sim t_k(r, \mathbf{0}, n(\mathbf{Z}'\mathbf{Z})^{-1}).$$

It is tempting to seek out a more flexible model via the Wishart distribution, but note if instead $\gamma_i | \boldsymbol{\Omega} \sim N_k(\mathbf{0}, \boldsymbol{\Omega})$, and $\boldsymbol{\Omega} \sim W_k^{-1}(r + k - 1, nr(\mathbf{Z}'\mathbf{Z})^{-1})$, the same marginal distribution is induced on γ_i. Here, $W_k^{-1}(r, \mathbf{R})$ is an inverted-Wishart distribution with the usual parameters $(r, \mathbf{R})$, $r > (k - 1)$. One can play around with different settings for various hyperparameters, but the end result is typically a multivariate t distribution or something close. For example, ref. [38] proposed a default random effects specification for generalized linear models; under the normal errors model their proposal is $\gamma_i | \boldsymbol{\Omega} \overset{iid}{\sim} N_k(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} | \sigma^2 \sim W^{-1}(k, k\mathbf{R})$, $\mathbf{R} = wc\sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1}$, where $w > 0$ is an inflation factor. Their induced marginal prior is $\gamma_i \sim t_k(1, \mathbf{0}, kwc\sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1})$. Note that our specification is not conditional on σ^2, otherwise, all these priors induce a multivariate t-distribution with similar covariance

structures. Ref. [38] compare their approach to the approximate uniform shrinkage prior of [39]. Ref. [40] extended half-t prior [33] to the multivariate setting so that the prior on the covariance matrix induces half-t priors on standard deviations and uniform priors on correlations.

We proceed to build a prior that reflects the prior knowledge on the overall marginal mean m and variance v of the data y_{ij}. Under prior (20) or (21) along with (19), we have $m = E(y_{ij})$ as before and now

$$v = \text{var}(y_{ij}) = \text{var}(\mathbf{x}'_{ij}\beta) + \text{var}(\mathbf{z}'_{ij}\gamma_i) + \text{var}(\epsilon_{ij}) \approx g_1 p + g_2 k + \sigma^2.$$

Certainly $\sigma^2 \leq v$ and $g_2 k \leq v - \sigma^2$ are reasonable bounds. The following default specification enforces the mean and variance constraint of $y_{ij} \sim (m, v)$:

$$\begin{aligned}
\beta|g, \sigma^2 &\sim N_p\left(m\mathbf{e}_1, \tfrac{n}{p}(v - \sigma^2 - gk)(\mathbf{X}'\mathbf{X})^{-1}\right). \\
\gamma_i|g &\overset{iid}{\sim} N_k\left(\mathbf{0}, gn(\mathbf{Z}'\mathbf{Z})^{-1}\right) \\
g|\sigma^2 &\sim \text{gb}\left(a_1, b_1, \tfrac{v-\sigma^2}{k}\right) \\
\sigma^2 &\sim \text{gb}(a_2, b_2, v) \\
m &\sim N(m_0, v/k_m), \; v^{-1} \sim \Gamma(k_v, v_0 k_v).
\end{aligned}$$
$$(22)$$

A uniform prior on σ^2 is specified $a_2 = b_2 = 1$; a uniform prior on g obtains from $a_1 = b_1 = 1$. When covariates $\mathbf{z}_{ij}$ come from quite different subpopulations across clusters, we recommend replacing the prior on γ_i in (22) with $\gamma_i|g \overset{ind.}{\sim} N_k(\mathbf{0}, gn_i(\mathbf{Z}'_i\mathbf{Z}_i)^{-1})$. The proposed prior (22) enables easy computation of the approximate amount of variation explained by random effects (i.e., gk/v) and fixed effects (i.e., $(v - \sigma^2 - gk)/v$) relative to the total v.

The priors on the fixed and random effect portions of the model are tied together and correlated; this is necessary to *a priori* conserve marginal variance. Ref. [41] note that, although variance components are usually modeled independently in the prior, typically as inverse-gamma, uniform, or half-Cauchy, they are "*linked as they are components of the total variation in the response...*" and suggests modeling them jointly as we do here, though via generalized multivariate gamma or multivariate log-normal distributions.

3.3. Hyper-Prior Elicitation for (M, V) in Mixed Models

Our prior in (22) requires specifying hyperparameters m_0, v_0, k_m and k_v in the hyperprior for (m, v). Suppose the historical data are $\mathbf{y}_o = \{y_{oij}|i = 1, \ldots, c_o; j = 1, \ldots, n_{oi}\} \sim (m, v)$. Set $M = \sum_{i=1}^{c_o} n_{oi}$. We need to extract sensible hyper-parameter values m_0, v_0, k_m and k_v so that the hyper-prior for (m, v) in (22) is close to the true posterior of (m, v) based on the historical data. Assume that the historical data can be approximately fit by the one-way random ANOVA: $y_{oij} = m + \gamma_i + \epsilon_{ij}$, $\gamma_i \overset{iid}{\sim} N(0, \sigma_{or}^2)$, $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_o^2)$. Unbiased estimates for m, σ_{or}^2 and σ_o^2 can be obtained using restricted maximum likelihood (REML) via the R function lmer in package lme4 [42], denoted as $\hat{m}_o$, $\hat{\sigma}_{or}^2$ and $\hat{\sigma}_o^2$. Then $\hat{v}_o = \hat{\sigma}_o^2 + \hat{\sigma}_{or}^2$ is an unbiased estimate of v, and $\hat{\rho}_o = \hat{\sigma}_{or}^2/\hat{v}_o$ is an estimate of the intraclass correlation coefficient. Based on some simulation trials, we find that the following posterior distributions approximately hold

$$m|\mathbf{y}_o, \hat{\rho}_o, v \sim N(\hat{m}_o, \tfrac{v}{n_{om}}), \; v^{-1}|\mathbf{y}_o, \hat{\rho}_o \sim \Gamma(\tfrac{n_{ov}}{2}, \hat{v}_y \tfrac{n_{ov}}{2}),$$
$$(23)$$

where $n_{om} = M/(1 + \hat{\rho}_o/(n_{o\lambda} - 1))$ and $n_{ov} = M/(1 + \hat{\rho}_o^2/(n_{o\lambda} - 1))$ can be interpreted as the effective sizes to account for the intraclass dependency, where $n_{o\lambda} = c_o/\sum_i \frac{1}{n_{oi}}$. Simple simulations (not shown here) reveal that the posterior distributions in (23) often provide us empirical coverage probabilities for (m, v) around 0.95, and the confidence width for v

is much narrower than the methods proposed in [43]. Further investigation is needed to understand the reason behind this. Fortunately, we use this approximate posterior to only select a reasonable hyperprior for (m, v) but not for our actual posterior inference based on the current data.

If one believes that the historical data $\mathbf{y}_o$ come from the same population as the current observed response data $\mathbf{y} = \{y_{ij} | i = 1, \ldots, c; j = 1, \ldots, n_i\}$, it is reasonable to set $m_0 = \hat{m}_o$, $v_0 = \hat{v}_o$, $k_m = n_{om}$ and $k_v = n_{ov}/2$ in the hyperprior of (m, v) in (22). Setting lower values for k_m and k_v puts less weight on the historical data relative to the current data. If the historical data are not available, we recommend setting $m_0 = \hat{m}$, $v_0 = \hat{v}$, $k_m = 2$ and $k_v = 1$, where $\hat{m}$ and $\hat{v}$ are the REML estimates of (m, v) based on the current response data $\mathbf{y}$.

For the random effects one-way ANOVA model (16), the prior (22) reduces to $\beta | g, \sigma^2 \sim N(m, v - \sigma^2 - g)$, $\gamma_i \sim N(0, g)$, $g | \sigma^2 \sim \text{gb}(a_1, b_1, v - \sigma^2)$ and $\sigma^2 \sim \text{gb}(a_2, b_2, v)$. In addition, the prior information of $y_{ij} \sim (m, v)$ indicates $\sigma^2 + g = v$ which further leads to $\beta = m$. Therefore, it is easy to show that the prior (22) for the random effects one-way ANOVA model finally reduces to

$$
\begin{gathered}
\gamma_i \sim N(0, g), \ \tfrac{\sigma^2}{v} \sim \text{beta}(a_2, b_2), \\
\beta \sim N(m_0, v/k_m), \ v^{-1} \sim \Gamma(k_v, v_0 k_v).
\end{gathered}
\tag{24}
$$

If we set $a_2 = b_2 = 1$ and $k_m = k_v = 0$, the prior in (24) is equivalent to

$$
\gamma_i \sim N(0, g), \ \tfrac{\sigma^2}{\sigma^2 + g} \sim \text{uniform}(0, 1), \ p(\beta, \sigma^2) \propto \tfrac{1}{\sigma^2},
\tag{25}
$$

which is exactly the shrinkage prior considered in [35]. That is, our prior naturally reduces to a well-known shrinkage prior for the random one-way ANOVA when there is no history information available for (m, v).

3.4. Rats Data Example

In the rat data example from WinBUGS manual [44], 30 rats' weights (in kg) were measured weekly for five weeks. Let y_{ij} be the weight of the ith rat measured in week j and a_{ij} be the corresponding age, $i = 1, \ldots, 30$, $j = 1, \ldots, 5$. Consider the mixed model (18) with $\mathbf{x}'_{ij} = (1, a_{ij})$, $\mathbf{z}'_{ij} = (1, a_{ij})$ and $\gamma_i = (\gamma_{i1}, \gamma_{i2})' \overset{iid}{\sim} N_k(\mathbf{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \text{diag}(\sigma_{r1}^2, \sigma_{r2}^2)$. Typically vague priors are used, e.g., $\gamma_{i1} \sim N(0, \sigma_{r1}^2)$, $\gamma_{i2} \sim N(0, \sigma_{r2}^2)$, $\sigma_{r1}^2, \sigma_{r2}^2 \overset{iid}{\sim} \Gamma^{-1}(0.01, 0.01)$. The marginal mean and variance for the response y_{ij} under this prior can be estimated via Monte Carlo (MC) by simulating $\sigma_{r1}^2, \sigma_{r2}^2, \gamma_{i1}, \gamma_{i2}$, $\beta = (\beta_0, \beta_1)' \sim N_2(\mathbf{0}, 10^2 \mathbf{I}_2)$, and $y_{ij}^{(l)} \overset{iid}{\sim} N(\mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\gamma_i, \sigma^2)$, $i = 1, \ldots, 30$, $j = 1, \ldots, 5$, $l = 1, \ldots, 1000$, yielding 1000 datasets $\{y_{ij}^{(l)} : i = 1, \ldots, 30, j = 1, \ldots, 5\}$, where $\sigma^2 \sim \Gamma^{-1}(0.01, 0.01)$. The average prior sample mean (across the 1000 datasets) and prior sample variance are around 2.5×10^{162} and ∞ (as reported in R), respectively. These substantially differ from the observed sample mean and variance of $\bar{y} = 0.243$ and $s_y^2 = 0.004$. In contrast, a similar simulation under our proposed new g-prior in (22) with $a_1 = b_1 = a_2 = b_2 = 1$, $m_0 = 0.243$, $v_0 = 0.004$, $k_m = 2$ and $k_v = 1$ yields an average sample mean of 0.249 with MC standard deviation of 0.144 and an average sample variance of 0.024 with MC standard deviation of 0.120. That is, the inference under our prior focuses on a much smaller set of potential models around those that could have conceivably generated the observed marginal data. The posterior estimates for β_0, β_1, and σ^2 under our proposed new g-prior are 0.1073 (0.0051), 0.0062 (0.0002), and 0.00004 (0.000006), respectively, where the values in parentheses are posterior standard deviations. The commonly used vague priors specified above yield to similar estimates but with much higher posterior standard deviations: 0.1067 (0.0059), 0.0061 (0.0049), and 0.00006 (0.000009).

3.5. Model Fitting via Block MCMC

Although the previous section portions variability due to $\mathbf{X}\beta$ and $\mathbf{Z}_i\gamma_i$ separately, Ref. [45] note that updating $(\beta', \gamma')'$ in one large block virtually eliminates problematic MCMC mixing, as β and γ are often *highly* correlated in the posterior. An optimal approach considers the full model (18) jointly

$$\mathbf{y} = \begin{bmatrix} \mathbf{X} & \tilde{\mathbf{Z}} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \epsilon, \tag{26}$$

where $\tilde{\mathbf{Z}} = \text{block-diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c) \stackrel{def}{=} \text{block-diag}\{\mathbf{Z}_i | i = 1, \ldots, c\}$. Under the prior (22), the full conditional for $(\beta', \gamma')'$ is

$$\begin{bmatrix} \beta \\ \gamma \end{bmatrix} \Big| \mathbf{y}, \sigma^2, g \sim N_{p+ck}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n),$$

where

$$\boldsymbol{\mu}_n = \frac{1}{\sigma^2} \boldsymbol{\Sigma}_n \left\{ \begin{bmatrix} \mathbf{X}' \\ \tilde{\mathbf{Z}}' \end{bmatrix} \mathbf{y} + \begin{bmatrix} \frac{p\sigma^2}{n(v-\sigma^2-gk)} \mathbf{X}'\mathbf{X}m\mathbf{e}_i \\ \mathbf{0} \end{bmatrix} \right\}$$

$$\boldsymbol{\Sigma}_n = \sigma^2 \left\{ \begin{bmatrix} \left(1 + \frac{p\sigma^2}{n(v-\sigma^2-gk)}\right)\mathbf{X}'\mathbf{X} & \mathbf{X}'\tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}'\mathbf{X} & \text{block-diag}\left\{\mathbf{Z}_i'\mathbf{Z}_i + \frac{w_i\sigma^2}{gn}\mathbf{Z}'\mathbf{Z} | i = 1, \ldots, c\right\} \end{bmatrix} \right\}^{-1}.$$

The full conditionals for σ^2 and g do not correspond to any known distributions, so an adaptive Metropolis algorithm [46] can be used.

4. Simulation Study

In all simulation studies, for each MCMC run, 5000 scans were thinned from 20,000 after a burn-in period of 2000 iterations; convergence diagnostics deemed this more than adequate. We use posterior means as the point estimates for all parameters. R functions to implement linear and linear mixed models using the proposed priors are provided in Supplementary Materials.

4.1. Simulation I: Fixed Effects Model

Simulations were carried out to evaluate the proposed methodology and compare it to the benchmark prior, local empirical Bayes (EB) approach and a hyper-g prior considered in [12]. Data were generated from the Gaussian regression model

$$y_i = \beta'\mathbf{x}_i + \epsilon_i, \ \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \ i = 1, \ldots, n,$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ and $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ip})'$. Let $\mathbf{X}_c$ be the usual centered design matrix for $(x_{i2}, \ldots, x_{ip})'$. The benchmark and EB methods consider the following priors

$$(\beta_2, \ldots, \beta_p) \stackrel{def}{=} \beta^* \sim N_{p-1}(\mathbf{0}, g\sigma^2(\mathbf{X}_c'\mathbf{X}_c)^{-1})$$
$$\beta_1 \sim N(0, 10^{10}) \tag{27}$$
$$\sigma^2 \sim \Gamma^{-1}(0.001, 0.001),$$

where $g = \max\{n, (p-1)^2\}$ is set for the benchmark method and $g = \max\{0, R^2(n - p)/(1 - R^2)/p\}$ is used for the EB approach, where R^2 is the R-squared value under the considered model. The hyper-g is given by

$$
\begin{aligned}
\beta^* &\sim N_{p-1}(\mathbf{0}, g\sigma^2(\mathbf{X}_c'\mathbf{X}_c)^{-1}) \\
\beta_1 &\sim N(0, 10^{10}) \\
\sigma^2 &\sim \Gamma^{-1}(0.001, 0.001) \\
\frac{g}{1+g} &\sim \text{beta}(1, h/2 - 1),
\end{aligned}
\tag{28}
$$

where we set $h = 3$ in all simulations which is the same as the setting used in [12].

4.1.1. Parameter Estimation

First we evaluate the performance for estimating model parameters using various methods. We generated $(x_{i2}, \dots, x_{ip})' \overset{iid}{\sim} N_{p-1}(\mathbf{1}, \boldsymbol{\Sigma}_\rho)$, where $\boldsymbol{\Sigma}_\rho$ has diagonals being 1 and off-diagonals being ρ. We set $p = 3$, $\sigma^2 = 1$, $\rho = 0.9$, and $\beta = (0.3, 0.3, 0.3)'$, yielding R-squared values around 0.26. The true marginal mean and variance of y_i are given by $m_T = E(y_i) = \beta_1 + \beta_2 + \beta_3 = 0.9$ and $v_T = var(y_{ij}) = \sigma^2 + (\beta_2, \beta_3)\boldsymbol{\Sigma}_\rho(\beta_2, \beta_3)' = 1.342$, respectively. We implemented our proposed prior in (5) with $a = b = 1$.

To evaluate how historical data can improve the parameter estimation accuracy, we additionally generated y_{oi}s of size $M = 50$ in the same way as generating y_is and considered three settings of the hyperprior for (m, v): (V1) new-true, when infinite historical data available, $m_0 = m_T$, $v_0 = v_T$, $k_m = 10^{10}$ and $k_v = 10^{10}$, i.e., (m, v) is fixed at the truth (m_T, v_T); (V2) new-hist, when a small set of historical data available, $m_0 = \bar{y}_o$, $v_0 = s_{y_o}^2$, $k_m = M$ and $k_v = M/2$; (V3) new-none, when no historical data available, $m_0 = \bar{y}$, $v_0 = s_y^2$, $k_m = 2$ and $k_v = 1$.

Let $\boldsymbol{\theta}$ be a generic parameter and $\hat{\boldsymbol{\theta}}$ be an estimate. The mean squared error (MSE) for $\hat{\boldsymbol{\theta}}$ is defined as MSE $= \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_j (\hat{\theta}_j - \theta_j)^2$. The bias for $\hat{\boldsymbol{\theta}}$ is defined as $\sum_j (\hat{\theta}_j - \theta_j)$. Table 1 reports the average bias and MSE values and coverage probabilities with interval widths across 500 Monte Carlo (MC) replicates. When $n = 100$, our method without using history information (new-none) performs very similarly to the other three completing methods. When a little history information is available, our prior (new-hist) has significantly lower MSE values and reduced interval widths on estimating σ^2 without compromising the coverage probabilities; the performance for estimating β_js is also slightly better than other approaches. When the true information on (m, v) is available, the estimation performance under our prior (new-true) is further improved comparing to new-hist. Regarding the estimation bias, we can see that all informative priors lead to biased estimates with a general trend that higher informativeness of the prior leads to larger biases. As the sample size increases to $n = 500$, although our methods (new-hist and new-true) still outperform other priors, the differences become smaller.

Table 1. Simulation I: Average biases (MSEs), coverage probabilities (interval widths) across 500 MC replicates in the simulation study for parameter estimation. Here, new-true, new-hist and new-none corresponds to the three hyperprior versions (V1), (V2) and (V3), respectively.

	Bias (MSE)			Coverage (Width)			
Method	β_1	$(\beta_2, \beta_3)'$	σ^2	β_1	β_2	β_3	σ^2
			$n = 100$				
new-true	0.041 (0.0196)	−0.050 (0.0894)	−0.012 (0.0092)	0.94 (0.55)	0.97 (0.86)	0.96 (0.86)	0.98 (0.43)
new-hist	0.038 (0.0202)	−0.045 (0.0906)	−0.011 (0.0157)	0.95 (0.56)	0.98 (0.88)	0.97 (0.87)	0.96 (0.51)
new-none	0.022 (0.0205)	−0.031 (0.0948)	0.011 (0.0206)	0.95 (0.57)	0.98 (0.90)	0.97 (0.90)	0.95 (0.57)
benchmark	−0.012 (0.0194)	0.003 (0.1049)	0.006 (0.0205)	0.96 (0.57)	0.97 (0.91)	0.96 (0.92)	0.95 (0.57)
EB	0.018 (0.0212)	−0.027 (0.0961)	0.023 (0.0219)	0.94 (0.56)	0.97 (0.90)	0.96 (0.90)	0.95 (0.58)
hyper-g	0.037 (0.0231)	−0.047 (0.0906)	0.034 (0.0229)	0.95 (0.59)	0.97 (0.89)	0.97 (0.89)	0.95 (0.60)
			$n = 500$				
new-true	0.010 (0.0043)	−0.012 (0.0216)	0.004 (0.0034)	0.96 (0.25)	0.95 (0.40)	0.94 (0.40)	0.96 (0.24)
new-hist	0.010 (0.0044)	−0.012 (0.0217)	−0.001 (0.0038)	0.95 (0.25)	0.94 (0.40)	0.94 (0.40)	0.96 (0.24)
new-none	0.007 (0.0044)	−0.008 (0.0219)	0.003 (0.0041)	0.95 (0.25)	0.95 (0.40)	0.94 (0.40)	0.96 (0.25)
benchmark	0.000 (0.0043)	−0.002 (0.0223)	0.003 (0.0041)	0.95 (0.25)	0.94 (0.40)	0.94 (0.40)	0.96 (0.25)
EB	0.006 (0.0044)	−0.007 (0.0219)	0.006 (0.0042)	0.95 (0.25)	0.94 (0.40)	0.94 (0.40)	0.96 (0.25)
hyper-g	0.009 (0.0045)	−0.011 (0.0217)	0.008 (0.0042)	0.95 (0.26)	0.95 (0.40)	0.94 (0.40)	0.95 (0.25)

4.1.2. Variable Selection

For a given $p \geq 2$, we generated $(x_{i2}, \ldots, x_{ip})' \stackrel{def}{=} \mathbf{x}_i^*$ as follows: (i) simulate $\mathbf{x}_i^* \stackrel{iid}{\sim} N_{p-1}(\mathbf{1}, \boldsymbol{\Sigma}_\rho)$; (ii) set the even elements of $\mathbf{x}_i^*$ to be binary by setting them to 0 if less than 1 and to be 1 if greater than 1. We set $\rho = 0.7$, $p = 16$ and $\beta = (\beta_l', \mathbf{0}')$, where β_l is the first l elements of β for $l = 1, 2, 3, 4, 7, 10, 13, 16$. That is, among the $p = 16$ covariates (including the intercept), there are l of them having non-zero coefficients. For each given l, we generated $\sigma^2 \sim \text{beta}(400, 100)$ and $\beta_l \sim N_l(2\mathbf{e}_1, \frac{n}{l}(\mathbf{X}_l'\mathbf{X}_l)^{-1}(1 - \sigma^2))$, where $\mathbf{X}_l$ is the design matrix for $(1, x_{i1}, \ldots, x_{il})$. These settings yield R-squared values ranging from 0.11 to 0.30 for $l = 1, 2, 3, 4, 7, 10, 13, 16$. For our method, we additionally generated y_{oi}s of size $M = 50$ in the same way as generating y_is and considered the same three versions of the hyper prior for (m, v) as in Section 4.1.1: (V1) new-true; (V2) new-hist; (V3) new-none. To compare our methods to the benchmark, EB and hyper-g approaches, we considered the following three cases under each prior: (C1) implement the variable selection procedure and obtain OLS estimation using the selected model; (C2) obtain Bayesian estimation using the true model; (C3) obtain Bayesian estimation using the full model. Here (C1) is used to compare the pure variable selection performance, (C2) is used to compare the predictive performance under the true model, and (C3) is used to compare the overall predictive performance when the model contains noisy covariates. For all Bayesian methods, posterior means $\hat{\beta}$ were used for estimating β.

Table 2 reports the average values of $\|\mathbf{X}\beta - \mathbf{X}\hat{\beta}\|^2/n$ across 200 MC replicates with $n = 100$. When OLS is used for fitting the selected model, the three versions of our methods perform very similarly, indicating that the history information on (m, v) has little influence on variable selection accuracy. Comparing to EB and hyper-g priors, our methods perform slightly better when the true model size is small ($l \leq 2$), and perform very similarly when $l \geq 3$. The benchmark prior works much better when the true model size is less than or equal to 7, but performs much worse when the true model size increases. The reason is that the benchmark prior sets $g = \max(n, (p-1)^2) = 225$ which leads to a more flat prior on β. When Bayesian estimation is used under the true or full model and there is some history information available on (m, v), our methods (both new-hist and new-true) outperform the other methods, and the benchmark prior is the worst due to its large choice of g. The only case where new-hist and new-true do not perform better is when the full model is fit but the

null model is the truth i.e., $l = 1$ under (C3), for which more historical data (see new-true) will help. Even when we don't have any history information on (m, v), the results under (C2) and (C3) show that our method performs slightly better than other methods, especially when $l \geq 4$.

Table 2. Simulation I: Average $\|\mathbf{X}\beta - \mathbf{X}\hat{\beta}\|^2 / n$ values across 200 MC replicates in the simulation study for variable selection.

Method	Size = 1	Size = 2	Size = 3	Size = 4	Size = 7	Size = 10	Size = 13	Size = 16
	(C1) OLS estimation using the selected model							
new-true	0.064	0.076	0.079	0.095	0.108	0.116	0.130	0.138
new-hist	0.062	0.074	0.078	0.095	0.107	0.116	0.130	0.137
new-none	0.056	0.072	0.079	0.094	0.107	0.116	0.130	0.137
benchmark	0.025	0.046	0.052	0.071	0.100	0.126	0.145	0.159
EB	0.110	0.087	0.081	0.095	0.108	0.115	0.130	0.137
hyper-*g*	0.094	0.081	0.079	0.093	0.107	0.114	0.131	0.138
	(C2) Bayesian estimation using the true model							
new-true	0.000	0.012	0.020	0.029	0.042	0.058	0.069	0.078
new-hist	0.007	0.013	0.021	0.030	0.044	0.061	0.072	0.083
new-none	0.010	0.015	0.023	0.032	0.047	0.063	0.074	0.085
benchmark	0.010	0.016	0.024	0.034	0.057	0.079	0.103	0.127
EB	0.010	0.016	0.025	0.034	0.049	0.065	0.076	0.088
hyper-*g*	0.010	0.016	0.025	0.034	0.048	0.064	0.075	0.086
	(C3) Bayesian estimation using the full model							
new-true	0.017	0.050	0.060	0.069	0.072	0.078	0.079	0.078
new-hist	0.028	0.056	0.065	0.073	0.077	0.082	0.083	0.083
new-none	0.026	0.058	0.067	0.075	0.079	0.083	0.084	0.085
benchmark	0.154	0.132	0.126	0.131	0.131	0.127	0.129	0.127
EB	0.018	0.057	0.069	0.078	0.082	0.087	0.087	0.088
hyper-*g*	0.023	0.057	0.067	0.075	0.080	0.084	0.085	0.086

4.2. Simulation II: Random One-Way ANOVA

Data are generated from the random one-way ANOVA model

$$y_{ij} = \beta_1 + \gamma_i + \epsilon_{ij}, \ \gamma_i \overset{iid}{\sim} N(0, \sigma_r^2), \ \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \ i = 1, \ldots, c, \ j = 1, \ldots, n_i.$$

where we set $\beta_1 = 2$, $c = 10$, $\sigma^2 = 1$. In addition, we consider $\sigma_r^2 = 0.25$, 0.5 and $n_i \sim$ dis-unif $[10, 15]$, where dis-unif $[a, b]$ represents a discrete uniform distribution with support being all integers with $[a, b]$. The true marginal mean and variance of y_{ij} are given by $m_T = E(y_{ij}) = \beta_1 = 2$ and $v_T = var(y_{ij}) = \sigma^2 + \sigma_r^2$, respectively. We implement our proposed default prior in (22) with $a_1 = b_1 = a_2 = b_2 = 1$ and the hyper-prior settings recommended in Section 3.3. Then σ_r^2 can be estimated from the posterior samples of g.

We additionally generate $\{y_{oij} | i = 1, \ldots, 10; j = 1, \ldots, 10\}$ in the same way as generating y_{ij}s and consider three versions of the hyper prior for (m, v): (V1) new-true, $m_0 = m_T$, $v_0 = v_T$, $k_m = 10^{10}$ and $k_v = 10^{10}$; (V2) new-hist, $m_0 = \hat{m}_o$, $v_0 = \hat{v}_o$, $k_m = n_{om}$ and $k_v = n_{ov}/2$; (V3) new-none, $m_0 = \hat{m}$, $v_0 = \hat{v}$, $k_m = 2$ and $k_v = 1$; see Section 3.3 for the definitions of these hyper-parameters. We also compare our methods to the $\sigma_r \sim$ uniform$(0, 10^2)$ prior [33], the $\sigma_r^2 \sim$ uniform$(0, 10^4)$ prior [47], the $\sigma_r^2 \sim \Gamma^{-1}(0.001, 0.001)$ prior [47] and the $\sigma^2/(\sigma^2 + \sigma_r^2) \sim$ uniform$(0, 1)$ shrinkage prior [35]. For these alternative priors, the typical priors $N(0, 10^3)$ and $\Gamma^{-1}(0.001, 0.001)$ are used on β_1 and σ^2, respectively.

Table 3 reports the average bias and MSE values and coverage probabilities with interval widths across 500 MC replicates, where the coverage probabilities for γ_is are

defined as the average coverage across all γ_is for $i = 1, \ldots, c$. Our approach with new-hist or new-true has significantly lower MSE values and narrower interval widths for estimating all model parameters while maintaining coverage probabilities around the nominal level 95% than other methods in all cases. Even when history information on (m, v) is not available, our method with new-none still has much lower MSE values for estimating σ_r^2 and narrower confidence interval widths than all other priors. Note that the induced prior under new-true essentially assumes that the prior variance of β_1 is zero, so we didn't report the coverage probability for β_1 here.

Table 3. Simulation II: Average biases (MSEs), coverage probabilities (interval widths) across 500 MC replicates in the simulation study for random one-way ANOVA model.

	Bias (MSE)				Coverage (Width)			
Method	β_1	γ_i	σ^2	σ_r^2	β_1	γ_i	σ^2	σ_r^2
				$\sigma_r^2 = 0.25$, $n_i \sim$ sample$(10, 15)$				
new-true	−0.000 (0.000)	−0.015 (0.619)	−0.010 (0.007)	0.010 (0.007)	-	0.94 (0.94)	0.97 (0.34)	0.97 (0.34)
new-hist	−0.000 (0.018)	−0.0003 (0.711)	−0.0001 (0.012)	0.055 (0.019)	0.96 (0.55)	0.95 (1.05)	0.95 (0.46)	0.97 (0.57)
new-none	0.018 (0.035)	−0.145 (0.796)	0.002 (0.015)	0.094 (0.034)	0.96 (0.78)	0.95 (1.15)	0.96 (0.51)	0.98 (0.85)
unif σ_r	0.017 (0.035)	−0.139 (0.814)	0.019 (0.016)	0.140 (0.067)	0.97 (0.84)	0.95 (1.20)	0.95 (0.52)	0.96 (1.13)
unif σ_r^2	0.018 (0.035)	−0.145 (0.801)	0.015 (0.016)	0.267 (0.140)	0.98 (0.96)	0.98 (1.30)	0.95 (0.52)	0.95 (1.64)
gamma	0.017 (0.035)	−0.133 (0.846)	0.026 (0.017)	0.057 (0.040)	0.94 (0.77)	0.93 (1.12)	0.96 (0.53)	0.93 (0.88)
shrink	0.018 (0.035)	−0.142 (0.796)	0.007 (0.015)	0.125 (0.046)	0.97 (0.83)	0.96 (1.19)	0.96 (0.51)	0.98 (0.99)
				$\sigma_r^2 = 0.5$, $n_i \sim$ sample$(10, 15)$				
new-true	−0.0000 (0.000)	−0.0006 (0.685)	0.012 (0.011)	−0.0012 (0.011)	-	0.94 (1.01)	0.96 (0.43)	0.96 (0.43)
new-hist	−0.0000 (0.029)	0.003 (0.892)	0.010 (0.013)	0.056 (0.045)	0.96 (0.71)	0.95 (1.20)	0.96 (0.49)	0.95 (0.89)
new-none	0.024 (0.059)	−0.206 (1.092)	0.007 (0.015)	0.109 (0.079)	0.96 (0.98)	0.96 (1.34)	0.95 (0.51)	0.98 (1.32)
unif σ_r	0.023 (0.059)	−0.199 (1.099)	0.018 (0.016)	0.264 (0.211)	0.97 (1.14)	0.97 (1.46)	0.96 (0.52)	0.97 (2.02)
unif σ_r^2	0.023 (0.059)	−0.195 (1.092)	0.016 (0.016)	0.471 (0.434)	0.99 (1.27)	0.98 (1.57)	0.96 (0.52)	0.95 (2.90)
gamma	0.022 (0.059)	−0.185 (1.115)	0.021 (0.016)	0.135 (0.127)	0.96 (1.05)	0.96 (1.38)	0.95 (0.53)	0.95 (1.61)
shrink	0.024 (0.059)	−0.205 (1.092)	0.011 (0.015)	0.168 (0.112)	0.97 (1.07)	0.96 (1.41)	0.95 (0.52)	0.98 (1.57)

4.3. Simulation III: Random Intercept Model

Data were generated from the mixed model

$$y_{ij} = \beta' \mathbf{x}_{ij} + \gamma_i + \epsilon_{ij}, \ \gamma_i \overset{iid}{\sim} N(0, \sigma_r^2), \ \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \ i = 1, \ldots, c, \ j = 1, \ldots, n_i,$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$, $\mathbf{x}_{ij} = (1, x_{ij2}, \ldots, x_{ijp})'$ and $(x_{ij2}, \ldots, x_{ijp})' \overset{iid}{\sim} N_{p-1}(\mathbf{1}, \Sigma_\rho)$. We set $p = 3$, $\rho = 0.9$, $\beta = (0.5, 0.5, 0.5)$, $c = 10$, $\sigma^2 = 1$. In addition, we consider $\sigma_r^2 = 0.25, 0.5$ and $n_i \sim$ dis-unif $[10, 15]$. The true marginal mean and variance of y_{ij} are given by $m_T = E(y_{ij}) = \beta_1 + \beta_2 + \beta_3 = 1.5$ and $v_T = var(y_{ij}) = \sigma^2 + \sigma_r^2 + (\beta_2, \beta_3)\Sigma_\rho(\beta_2, \beta_3)' = 1.95 + \sigma_r^2$, respectively. The prior settings are the same as those used in Section 4.2.

Table 4 reports the average bias and MSE values and coverage probabilities with interval widths across 500 MC replicates. Our approach with new-hist or new-true has significantly lower MSE values and narrower interval widths for estimating all model parameters while maintaining coverage probabilities around the nominal level 95% than other methods in all cases. Even when history information on (m, v) is not available, our method with new-none still has much lower MSE values for estimating (β_2, β_3) and σ_r^2 with slightly narrower confidence interval widths than all other priors.

Table 4. Simulation III: Average biases (MSEs), coverage probabilities (interval widths) across 500 MC replicates in the simulation study for random intercept model.

Method	Bias (MSE)					Coverage (Width)					
	β_1	(β_2, β_3)	γ_i	σ^2	σ_r^2	β_1	β_2	β_3	γ_i	σ^2	σ_r^2
	$\sigma_r^2 = 0.25$, $n_i \sim \text{sample}(10, 15)$										
new-true	0.025 (0.034)	−0.0032 (0.085)	0.053 (0.772)	0.038 (0.017)	0.084 (0.022)	0.98 (0.81)	0.95 (0.82)	0.95 (0.82)	0.96 (1.15)	0.96 (0.52)	0.98 (0.64)
new-hist	0.028 (0.035)	−0.0034 (0.085)	0.046 (0.787)	0.029 (0.017)	0.083 (0.026)	0.97 (0.82)	0.94 (0.81)	0.94 (0.81)	0.96 (1.14)	0.96 (0.52)	0.98 (0.70)
new-none	0.025 (0.043)	−0.0034 (0.085)	0.076 (0.828)	0.027 (0.017)	0.106 (0.039)	0.94 (0.83)	0.94 (0.81)	0.94 (0.81)	0.95 (1.15)	0.97 (0.53)	0.97 (0.88)
unif σ_r	−0.0011 (0.042)	0.002 (0.090)	0.074 (0.846)	0.027 (0.018)	0.154 (0.075)	0.96 (0.93)	0.94 (0.83)	0.94 (0.83)	0.96 (1.22)	0.96 (0.53)	0.96 (1.15)
unif σ_r^2	−0.0011 (0.041)	0.002 (0.090)	0.072 (0.832)	0.022 (0.017)	0.273 (0.146)	0.98 (1.03)	0.94 (0.82)	0.95 (0.83)	0.97 (1.31)	0.97 (0.53)	0.93 (1.55)
gamma	−0.0010 (0.042)	0.001 (0.090)	0.075 (0.941)	0.052 (0.022)	0.039 (0.043)	0.94 (0.84)	0.94 (0.84)	0.95 (0.84)	0.92 (1.12)	0.95 (0.57)	0.96 (0.94)
shrink	−0.0010 (0.041)	0.001 (0.090)	0.072 (0.827)	0.015 (0.016)	0.138 (0.052)	0.96 (0.93)	0.94 (0.82)	0.94 (0.82)	0.96 (1.21)	0.97 (0.52)	0.97 (1.00)
	$\sigma_r^2 = 0.5$, $n_i \sim \text{sample}(10, 15)$										
new-true	0.023 (0.046)	−0.0032 (0.086)	0.071 (0.976)	0.035 (0.017)	0.055 (0.028)	0.97 (0.94)	0.94 (0.82)	0.95 (0.82)	0.96 (1.28)	0.96 (0.52)	0.99 (0.82)
new-hist	0.025 (0.050)	−0.0032 (0.086)	0.061 (1.018)	0.028 (0.017)	0.069 (0.048)	0.97 (0.96)	0.94 (0.81)	0.94 (0.82)	0.96 (1.29)	0.96 (0.52)	0.98 (1.00)
new-none	0.019 (0.066)	−0.0031 (0.086)	0.109 (1.129)	0.029 (0.018)	0.131 (0.090)	0.93 (0.99)	0.94 (0.82)	0.95 (0.82)	0.95 (1.31)	0.96 (0.53)	0.96 (1.38)
unif σ_r	−0.0014 (0.065)	0.002 (0.091)	0.107 (1.134)	0.025 (0.018)	0.292 (0.238)	0.97 (1.21)	0.94 (0.83)	0.95 (0.83)	0.97 (1.48)	0.97 (0.53)	0.95 (2.07)
unif σ_r^2	−0.0015 (0.065)	0.002 (0.091)	0.107 (1.127)	0.023 (0.017)	0.481 (0.445)	0.98 (1.33)	0.94 (0.83)	0.94 (0.83)	0.98 (1.59)	0.97 (0.53)	0.93 (2.73)
gamma	−0.0014 (0.066)	0.001 (0.091)	0.106 (1.208)	0.042 (0.021)	0.133 (0.145)	0.94 (1.11)	0.94 (0.84)	0.95 (0.84)	0.95 (1.40)	0.95 (0.57)	0.96 (1.74)
shrink	−0.0013 (0.065)	0.002 (0.091)	0.102 (1.127)	0.019 (0.017)	0.192 (0.126)	0.96 (1.15)	0.94 (0.83)	0.95 (0.83)	0.96 (1.43)	0.97 (0.53)	0.97 (1.60)

4.4. Simulation IV: Linear Mixed Model

Data were generated from the mixed model

$$y_{ij} = \beta' \mathbf{x}_{ij} + \gamma_i' \mathbf{z}_{ij} + \epsilon_{ij}, \ \gamma_i \overset{iid}{\sim} N_k(\mathbf{0}, \boldsymbol{\Omega}_i), \ \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2), \ i = 1, \dots, c, \ j = 1, \dots, n_i,$$

where $\beta = (\beta_1, \beta_2, \beta_3)' = (0.5, 0.5, 0.5)'$, $\mathbf{x}_{ij} = (1, x_{ij2}, x_{ij3})' \overset{ind}{\sim} N_3(\mathbf{1}, \boldsymbol{\Sigma}_i)$, $k = 2$, $\gamma_i = (\gamma_{i1}, \gamma_{i2})'$, $\mathbf{z}_{ij} = (1, x_{ij2})'$, and $\boldsymbol{\Omega}_i = g(\mathbf{11}' + \boldsymbol{\Sigma}_i)^{-1}_{[1:2,1:2]}$. Here we set $\sigma^2 = 1$, $c = 10$, $\boldsymbol{\Sigma}_i = \text{diag}(0, 1, 1)$ for $i = 1, \dots, 5$ and $\boldsymbol{\Sigma}_i = \text{diag}(0, 4, 4)$ for $i = 6, \dots, 10$. Under this setting, we have the total random effective variance equal to $\text{var}(\mathbf{z}_{ij}' \gamma) = gk = 2g \overset{def}{=} \sigma_r^2$. We consider $\sigma_r^2 = 0.5, 1$ and $n_i \sim \text{dis-unif}\,[10, 15]$.

We implement our proposed default prior in (22) with $a_1 = b_1 = a_2 = b_2 = 1$ and a more general version of it with $\gamma_i | g \overset{ind.}{\sim} N_k(\mathbf{0}, gn_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1})$ (denoted as new-i below). Regarding the hyperprior of (m, v), we only consider new-hist and new-none as defined in Section 4.2, considering that the true marginal mean and variance of y_{ij} are not available in closed forms. We then compare our methods to the prior proposed in [38]: $\gamma_i | \boldsymbol{\Omega} \overset{iid}{\sim} N_k(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} | \sigma^2 \sim W^{-1}(k, k\mathbf{R})$, $\mathbf{R} = c\sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}$.

Table 5 reports the average bias and MSE values and coverage probabilities with interval widths across 500 MC replicates, where the coverage probabilities for γ_{ij}'s are defined as the average coverage across all γ_{ij}'s over $i = 1, \dots, c$ for each $j = 1, 2$. Comparing our default prior in (22) with its more general version new-i, the new-i method has lower MSE values for estimating most model parameters and is markedly better for estimating γ_{ij} and σ_r^2. Comparing to our default prior (22) and the prior in [38] (both assuming homogeneous covariance for γ_i), our prior has much lower MSE values and narrower interval widths for estimating $\gamma_i j$s while maintaining coverage probabilities around the nominal level 95%. When the more general prior new-i is used, our method consistently performs better than [38] on estimating all model parameters.

Table 5. Simulation IV: Average biases (MSEs), coverage probabilities (interval widths) across 500 MC replicates in the simulation study for general mixed model. Here, the suffix -i refers to the prior in (22) with $\gamma_i | g \overset{ind.}{\sim} N_k\left(\mathbf{0}, g\frac{n_i}{w_i}(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\right)$; KN refers to the prior introduced in [38].

Method	Bias (MSE)					Coverage (Width)						
	β_1	(β_2, β_3)	γ_i	σ^2	σ_r^2	β_1	β_2	β_3	γ_{1i}	γ_{2i}	σ^2	σ_r^2
$\sigma_r^2 = 0.5$, $n_i \sim \text{sample}(10, 15)$												
new-hist	0.084 (0.048)	−0.0071 (0.020)	−0.160 (2.078)	0.037 (0.019)	0.093 (0.063)	0.96 (0.90)	0.94 (0.49)	0.96 (0.23)	0.93 (1.40)	0.91 (0.80)	0.95 (0.55)	0.95 (0.98)
new-hist-i	0.077 (0.049)	−0.0062 (0.018)	−0.184 (2.045)	0.030 (0.018)	0.038 (0.043)	0.94 (0.86)	0.92 (0.45)	0.96 (0.23)	0.94 (1.46)	0.94 (0.90)	0.96 (0.55)	0.96 (0.87)
new-none	0.086 (0.056)	−0.0069 (0.021)	−0.189 (2.129)	0.035 (0.019)	0.119 (0.091)	0.94 (0.92)	0.94 (0.50)	0.95 (0.23)	0.93 (1.41)	0.91 (0.81)	0.95 (0.56)	0.94 (1.16)
new-none-i	0.081 (0.057)	−0.0063 (0.019)	−0.211 (2.101)	0.026 (0.019)	0.045 (0.053)	0.92 (0.87)	0.91 (0.46)	0.96 (0.23)	0.94 (1.47)	0.94 (0.90)	0.96 (0.55)	0.96 (0.97)
KN	0.010 (0.050)	0.005 (0.019)	−0.167 (2.209)	0.046 (0.019)	-	0.96 (0.95)	0.95 (0.53)	0.96 (0.23)	0.93 (1.40)	0.91 (0.81)	0.96 (0.55)	-
$\sigma_r^2 = 1$, $n_i \sim \text{sample}(10, 15)$												
new-hist	0.091 (0.074)	−0.103 (0.035)	0.096 (2.735)	0.019 (0.019)	0.125 (0.165)	0.96 (1.11)	0.89 (0.60)	0.95 (0.23)	0.94 (1.62)	0.90 (0.94)	0.94 (0.55)	0.93 (1.47)
new-hist-i	0.080 (0.070)	−0.0089 (0.028)	0.068 (2.544)	0.013 (0.019)	0.010 (0.114)	0.95 (1.06)	0.91 (0.57)	0.95 (0.23)	0.95 (1.66)	0.94 (1.01)	0.94 (0.54)	0.95 (1.32)
new-none	0.086 (0.091)	−0.0096 (0.036)	0.083 (2.848)	0.021 (0.020)	0.240 (0.330)	0.94 (1.16)	0.90 (0.63)	0.95 (0.23)	0.94 (1.66)	0.91 (0.97)	0.94 (0.55)	0.91 (1.98)
new-none-i	0.080 (0.087)	−0.0088 (0.029)	0.055 (2.670)	0.013 (0.019)	0.057 (0.172)	0.92 (1.09)	0.90 (0.58)	0.95 (0.23)	0.94 (1.67)	0.94 (1.02)	0.94 (0.55)	0.92 (1.61)
KN	−0.0014 (0.088)	0.002 (0.035)	0.104 (3.050)	0.053 (0.023)	-	0.95 (1.23)	0.93 (0.70)	0.95 (0.23)	0.93 (1.68)	0.91 (1.00)	0.94 (0.58)	-

5. Discussion

Prior elicitation plays an important role in Bayesian inference. We have proposed a novel, yet remarkably simple class of informative g-priors for linear mixed models elicited from existing information on the marginal distribution of the responses. The prior is firstly developed for the linear regression model (2) assuming that a subject-matter expert has information on the marginal distribution $y_i \overset{iid}{\sim} (m, v)$. A simple, intuitive interpretation of the prior is obtained: when $\sigma^2 = v$ the model explains nothing (i.e., reduces to the null model), when $\sigma^2 = 0$ the model explains all variability in responses; furthermore, the use of a generalized beta prior on $\sigma^2 \in [0, v]$ allows one to specify the prior information on the amount of variation explained by the considered model. The proposed prior also naturally reduces to a modified version of the hyper-g/n prior introduced in [12] when there is no history information available for (m, v). Under the Gaussian linear regression models with the proposed g-prior, Bayes factors for comparing all possible submodels can be easily computed for the purpose of variable selection and do not encounter the information paradox commonly seen in Zellner's g-priors with fixed g. Our approach is further extended for use in linear mixed models. Interesting relationships between the proposed g-priors and some other commonly used priors in mixed models are discussed. For example, under the random effect one-way ANOVA, the proposed prior (22) with a reference hyper prior on (m, v) reduces exactly to the shrinkage prior of [35]. Posterior sampling for all considered models can be obtained using JAGS via R. Finally, extensive simulation studies reveal that the proposed g-prior outperforms almost all other approaches under consideration when some history information on (m, v) is available. Even without historical data, better performance of the proposed new g-prior over other priors is still seen in many settings. Interesting generalizations of the proposed idea may include additive penalized B-spline regression, variable selection in the linear mixed models and prior elicitation for generalized linear mixed models. Recently, Ref. [48] proposed two informative priors for the between-cluster slope in a multilevel latent covariate model. However, extension of their methods to multiple covariates has not been investigated. It would be interesting to extend the proposed g-prior here to general multilevel latent covariate models.

References

1. Sun, C.Q.; Prajna, N.V.; Krishnan, T.; Mascarenhas, J.; Rajaraman, R.; Srinivasan, M.; Raghavan, A.; O'Brien, K.S.; Ray, K.J.; McLeod, S.D.; et al. Expert Prior Elicitation and Bayesian Analysis of the Mycotic Ulcer Treatment Trial I. *Investig. Ophthalmol. Vis. Sci.* **2013**, *54*, 4167–4173. [CrossRef] [PubMed]

2. Hampson, L.V.; Whitehead, J.; Eleftheriou, D.; Brogan, P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Stat. Med.* **2014**, *33*, 4186–4201. [CrossRef]

3. Zhang, G.; Thai, V.V. Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review. *Saf. Sci.* **2016**, *87*, 53–62. [CrossRef]

4. Food and Drug Administration. Guidance for the use of Bayesian statistics in medical device clinical trials. *Guid. Ind. Fda Staff.* **2010**, *2006*, 1–50.

5. O'Hagan, A. Eliciting expert beliefs in substantial practical applications. *J. R. Stat. Soc. Ser.* **1998**, *47*, 21–35.

6. Kinnersley, N.; Day, S. Structured approach to the elicitation of expert beliefs for a Bayesian-designed clinical trial: A case study. *Pharm. Stat.* **2013**, *12*, 104–113. [CrossRef]

7. Dallow, N.; Best, N.; Montague, T.H. Better decision making in drug development through adoption of formal prior elicitation. *Pharm. Stat.* **2018**, *17*, 301—316. [CrossRef]

8. Hartmann, M.; Agiashvili, G.; Bürkner, P.; Klami, A. Flexible Prior Elicitation via the Prior Predictive Distribution. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*; Peters, J., Sontag, D., Eds.; PMLR: London, UK, 2020; Volume 124, pp. 1129–1138.

9. Zellner, A. Applications of Bayesian Analysis in Econometrics. *Statistician* **1983**, *32*, 23–34. [CrossRef]

10. Zellner, A. On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions. In *Proceedings of the Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*; North-Holland/Elsevier: Amsterdam, The Netherlands, 1986; pp. 233–243.

11. Li, Y.; Clyde, M.A. Mixtures of *g*-priors in generalized linear models. *J. Am. Stat. Assoc.* **2018**, *113*, 1828–1845. [CrossRef]

12. Liang, F.; Paulo, R.; Molina, G.; Clyde, M.A.; Berger, J.O. Mixtures of *g* priors for Bayesian variable selection. *J. Am. Stat. Assoc.* **2008**, *103*, 410–423. [CrossRef]

13. Bedrick, E.J.; Christensen, R.; Johnson, W. A New Perspective on Priors for Generalized Linear Models. *J. Am. Stat. Assoc.* **1996**, *91*, 1450–1460. [CrossRef]

14. Hosack, G.R.; Hayes, K.R.; Barry, S.C. Prior elicitation for Bayesian generalised linear models with application to risk control option assessment. *Reliab. Eng. Syst. Saf.* **2017**, *167*, 351–361. [CrossRef]

15. Ibrahim, J.G.; Chen, M.H. Power prior distributions for regression models. *Stat. Sci.* **2000**, *15*, 46–60.

16. Ibrahim, J.G.; Chen, M.H.; Sinha, D. On optimality properties of the power prior. *J. Am. Stat. Assoc.* **2003**, *98*, 204–213. [CrossRef]

17. Hobbs, B.P.; Carlin, B.P.; Mandrekar, S.J.; Sargent, D.J. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **2011**, *67*, 1047–1056. [CrossRef] [PubMed]

18. Ibrahim, J.G.; Chen, M.H.; Gwon, Y.; Chen, F. The power prior: Theory and applications. *Stat. Med.* **2015**, *34*, 3724–3749. [CrossRef] [PubMed]

19. Agliari, A.; Parisetti, C.C. A-*g* Reference Informative Prior: A Note on Zellner's g-Prior. *J. R. Stat. Soc. Ser. D* **1988**, *37*, 271–275. [CrossRef]

20. van Zwet, E. A default prior for regression coefficients. *Stat. Methods Med. Res.* **2019**, *28*, 3799–3807. [CrossRef]

21. Plummer, M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria, 20–22 March 2003; Hornik, K., Leisch, F., Zeileis, A., Eds.; ISSN 1609-395X.

22. Su, Y.S.; Yajima, M. *R2jags: Using R to Run 'JAGS'*; R Package Version 0.5-7; R Foundation for Statistical Computing: Vienna, Austria, 2015.

23. Hanson, T.E.; Branscum, A.J.; Johnson, W.O. Informative *g*-Priors for Logistic Regression. *Bayesian Anal.* **2014**, *9*, 597–612. [CrossRef]

24. Lally, N.R. The Informative g-Prior vs. Common Reference Priors for Binomial Regression with an Application to Hurricane Electrical Utility Asset Damage Prediction. Master's Thesis, University of Connecticut, Mansfield, CT, USA, 31 July 2015 .

25. Carlin, B.P.; Gelfand, A.E. An iterative Monte Carlo method for nonconjugate Bayesian analysis. *Stat. Comput.* **1991**, *1*, 119–128. [CrossRef]

26. Liu, C.; Martin, R.; Syring, N. Efficient simulation from a gamma distribution with small shape parameter. *Comput. Stat.* **2017**, *32*, 1767–1775. [CrossRef]

27. Gabry, J.; Simpson, D.; Vehtari, A.; Betancourt, M.; Gelman, A. Visualization in Bayesian workflow. *J. R. Stat. Soc. Ser. A* **2019**, *182*, 389–402. [CrossRef]

28. Gelman, A.; Simpson, D.; Betancourt, M. The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy* **2017**, *19*, 555. [CrossRef]

29. Wesner, J.S.; Pomeranz, J.P.F. Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. *Ecosphere* **2021**, *12*, e03739. [CrossRef]

30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

31. Murphy, K.P. *Conjugate Bayesian Analysis of the Gaussian Distribution*; Technical Report; University of British Columbia: Vancouver, BC, Canada, 3 October 2007.

32. Berger, J.O.; Pericchi, L.R.; Ghosh, J.; Samanta, T.; De Santis, F.; Berger, J.; Pericchi, L. Objective Bayesian methods for model selection: Introduction and comparison. *Lect.-Notes-Monogr. Ser.* **2001**, 135–207.

33. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **2006**, *1*, 515–533. [CrossRef]
34. Box, G.E.P.; Tiao, G.C. *Bayesian Inference in Statistical Analysis*; Addison-Wesley: Reading, MA, USA, 1973.
35. Daniels, M.J. A prior for the variance in hierarchical models. *Can. J. Stat.* **1999**, *27*, 567–578. [CrossRef]
36. Wang, M. Mixture of *g*-priors for analysis of variance models with a divergining number of parameters. *Bayesian Anal.* **2017**, *12*, 511–532. [CrossRef]
37. Lin, P.E. Some characterizations of the multivariate t distribution. *J. Multivar. Anal.* **1972**, *2*, 339–344. [CrossRef]
38. Kass, R.E.; Natarajan, R. A default conjugate prior for variance components in a generalized linear mixed models (Comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 535–542. [CrossRef]
39. Natarajan, R.; Kass, R.E. Reference Bayesian methods for generalized linear mixed models. *J. Am. Stat. Assoc.* **2000**, *95*, 227–237. [CrossRef]
40. Huang, A.; Wand, M.P. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* **2013**, *8*, 439–452. [CrossRef]
41. Demirhan, H.; Kalaylioglu, Z. Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models. *J. Multivar. Anal.* **2015**, *135*, 163–174. [CrossRef]
42. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
43. Burdick, R.K.; Borror, C.M.; Montgomery, D.C. *Design and Analysis of Gauge R and R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*; SIAM: Philadelphia, PA, USA, 2005.
44. Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. *WinBUGS User Manual, Version 1.4*; Medical Research Council Biostatistics Unit: Cambridge, UK, 2003.
45. Sargent, D.J.; Hodges, J.S.; Carlin, B.P. Structured Markov Chain Monte Carlo. *J. Comput. Graph. Stat.* **2000**, *9*, 217–234.
46. Haario, H.; Saksman, E.; Tamminen, J. An Adaptive Metropolis Algorithm. *Bernoulli* **2001**, *7*, 223–242. [CrossRef]
47. Brown, W.J.; Draper, D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal.* **2006**, *1*, 473–514. [CrossRef]
48. Zitzmann, S.; Helm, C.; Hecht, M. Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: A comparative investigation of two different approaches. *Front. Psychol.* **2021**, *11*, 611267. [CrossRef]