Scientific Research Publishing

Open Journal of Modelling and Simulation, 2022, 10, 391-416
https://www.scirp.org/journal/ojmsi
ISSN Online: 2327-4026
ISSN Print: 2327-4018

Bayesian Variable Selection for Mixture Process Variable Design Experiment

Sadiah M. A. Aljeddani

Department of Mathematics, Al-Lith College, Umm Al-Qura University, Al-Lith, Saudi Arabia
Email: smgadany@uqu.edu.sa

How to cite this paper: Aljeddani, S.M.A. (2022) Bayesian Variable Selection for Mixture Process Variable Design Experiment. *Open Journal of Modelling and Simulation*, **10**, 391-416.
https://doi.org/10.4236/ojmsi.2022.104022

Received: September 8, 2022
Accepted: October 15, 2022
Published: October 18, 2022

Abstract

This paper discussed Bayesian variable selection methods for models from split-plot mixture designs using samples from Metropolis-Hastings within the Gibbs sampling algorithm. Bayesian variable selection is easy to implement due to the improvement in computing via MCMC sampling. We described the Bayesian methodology by introducing the Bayesian framework, and explaining Markov Chain Monte Carlo (MCMC) sampling. The Metropolis-Hastings within Gibbs sampling was used to draw dependent samples from the full conditional distributions which were explained. In mixture experiments with process variables, the response depends not only on the proportions of the mixture components but also on the effects of the process variables. In many such mixture-process variable experiments, constraints such as time or cost prohibit the selection of treatments completely at random. In these situations, restrictions on the randomisation force the level combinations of one group of factors to be fixed and the combinations of the other group of factors are run. Then a new level of the first-factor group is set and combinations of the other factors are run. We discussed the computational algorithm for the Stochastic Search Variable Selection (SSVS) in linear mixed models. We extended the computational algorithm of SSVS to fit models from split-plot mixture design by introducing the algorithm of the Stochastic Search Variable Selection for Split-plot Design (SSVS-SPD). The motivation of this extension is that we have two different levels of the experimental units, one for the whole plots and the other for subplots in the split-plot mixture design.

Keywords

Variable Selection, Bayesian Analysis, Mixture Experiment, Split-Plot Design

1. Introduction

Bayesian methods are important approaches due to their ability to quantify uncertainty. In such an approach, prior distributions that represent subjective be-

liefs about parameters are assigned to the regression coefficients. By applying Bayes' rule, prior beliefs are updated by the data and transformed into posterior distributions, on which all inference is based.

An important task in regression building is to determine which variables should be included in the model. Therefore, the principle of Bayesian variable selection is to get the large and the small effects distinctive, and effective prior essences mass around zero and distribute the remaining over the parameter space. Such a prior represents the fact that there are small coefficients close to zero on one hand and larger coefficients on the other hand. These priors can be built as a combination of two distributions, along with a narrow normal continuous distribution centred at zero with a small variance called a "spike", and the other with a flat normal continuous distribution with a large variance to spread over a wider range of parameter values called a "slab". This type of priors is called "Spike-and-Slab" priors (see Figure 1). Practically saying, those priors are beneficial for purposes of variable selection because they permit the classification of the regression coefficients into two groups: one group with large, significant effects, and the other group with small, negligible effects.

As reviewed by [1] [2] introduced Bayesian variable selection via "spike-and-slab" prior distributions. The spike prior that they used was a probability mass at zero to remove the non-significant variables. Their slab is the uniform distribution with a large symmetric range in order to keep the significant variables. Following their work, many priors were proposed to implement the spike-and-slab property. [3] proposed the Stochastic Search Variable Selection (SSVS) in which the coefficients are sampled from a mixture of two normal distributions with different variances. The spike part is the distribution with a small variance while the slab

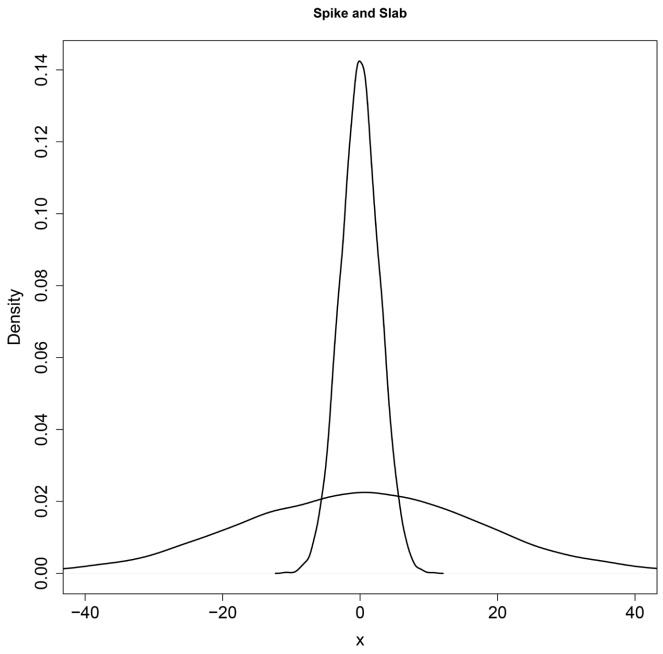

Figure 1. Gaussian mixture prior for β.

part is the distribution with a much larger variance. Also, [4] proposed positive mass at zero for the spike part and a normal distribution for the slab part. In addition, [5] proposed a Bayesian approach for model selection in fractionated split-plot experiments with application to "robust-parameter design". In their work, they extend the SSVS algorithm of [6] to account for the split-plot error structure. They derive an expression for the posterior probability of a model that requires the computation of, at most, two unidimensional integrals, and employ this quantity for model selection. They were able to integrate the coefficients and the variance components from the joint posterior distribution of all parameters because they use the conjugate normal-inverse gamma prior for these parameters. The integrals are computed with Gaussian quadrature, and Global and Local search algorithms to find models with high posterior probabilities.

The above review summarised the former research and analyzed the problems in the Bayesian variable selection field. On the other side, we notice that the frequentist methods analysis of this kind of experimental design would fail in high Type I error rate [7]. The difficulty of using the frequentist analysis to estimate the variance components is explained in [8]. To overcome this issue, we apply the Bayesian variable selection as this method introduces prior distribution about the variance component in the linear mixed models.

The novel contribution of this paper to Bayesian variable selection is motivated by a very specific experimental design of data from experiments subject to restricted randomisation. We have two different levels of the experimental units one for the whole plots and the other for the subplots in the split-plot design; see, for example, [9]. To address this issue, we adapt the SSVS algorithm in which we sample the subplot coefficients using a mixture of normal posterior distributions with a slab variance different from the slab variance which will be used in the mixture of normal posterior distributions for the whole-plot coefficients. This method reduces Type I and II error rates as well as reduces the prediction error for split-plot design rather than applying the SSVS algorithm in which all coefficients will be sampled from a mixture of normal posterior distributions with one slab variance. We called the approached method the Stochastic Search Variable Selection for Split-Plot Design (SSVS-SPD). The frequentist analysis is dependent on the estimates of the variance components, yet these estimates cannot be precisely calculated because of the deficiency of the degrees of freedom for the random effects in the split-plot design. This issue was discussed by [8]. Introducing a prior distribution for the variance components in the linear mixed model provides additional information to overcome the problem of variance estimation.

This paper differs from [10] in which the dataset used in this paper is the Vinyl Thickness experiment, which had not been used in variable selection. Also, it differs in the choice of the shape parameter of the correlation parameter as in this paper we used $a = b = 2$ to fit this data. A preprint has previously been published [11].

2. Split-Plot Design and Sample Model

The model for the block experiments includes two types of errors: block error and residual error. Hence, linear mixed models (LMMs) are used to analyse responses from the blocked experiments.

Linear mixed-effects models (LMMs) introduce correlations between observations using random effects. This leads to the use of generalised least squares (GLS) estimation, combined with restricted maximum likelihood estimation (REML) of the variance components as will be discussed. This type of analysis is used by the most design of experiments textbooks that deal with blocked designs. In matrix notation, the model corresponding to a blocked design is written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y}$ is $n\times 1$ vector of observations on the response of interest, $\mathbf{X}$ is the $n\times p$ model design matrix containing the polynomial expansions of the m factor levels at the n experimental runs, $\boldsymbol{\beta}$ is the $p\times 1$ vector of unknown fixed parameters, $\mathbf{Z}$ is an $n\times b$ random design matrix which represents the allocation of the runs to blocks, and whose $(i,j)^{\text{th}}$ element is one where the i^{th} observation belongs to the j^{th} blocks, and zero otherwise. If the runs of the experiment are grouped per block, then $\mathbf{Z}$ is of the form

$$\mathbf{Z} = \operatorname{diag}\left[\mathbf{1}_{k_1}, \mathbf{1}_{k_2}, \cdots, \mathbf{1}_{k_b}\right], \tag{2}$$

where $\mathbf{1}_k$ is a k vector of ones, and $k_1, k_2, \cdots, k_b$ are the blocks sizes. The random effects of the b blocks are contained within the $b\times 1$ vector $\boldsymbol{\gamma}$, and the random errors are contained within the $n\times 1$ vector $\boldsymbol{\varepsilon}$. It is assumed that $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are independent and normally distributed, *i.e.* $\operatorname{Cov}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = \mathbf{0}_{b\times n}$, where $\mathbf{0}_{b\times n}$ is the $b\times n$ matrix of zeros. Hence, $\boldsymbol{\gamma} \sim \mathrm{N}\left(\mathbf{0}_b, \sigma_\gamma^2 \mathbf{I}_b\right)$, and $\boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{0}_n, \sigma_\varepsilon^2 \mathbf{I}_n\right)$, where $\mathbf{0}_b$ and $\mathbf{0}_n$ are the b and n column vectors of zeros respectively, and $\mathbf{I}_b$ and $\mathbf{I}_n$ are the b-dimensional and n-dimensional identity matrices respectively.

Under these assumptions, $\mathbf{Y}$ is a normally distributed random variable with mean $\mathbb{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, and the variance-covariance matrix of the response $\mathbf{Y}$ can be written as

$$\mathbf{V} = \operatorname{Var}(\mathbf{Y}) = \operatorname{Var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \tag{3}$$

$$= \operatorname{Var}(\mathbf{Z}\boldsymbol{\gamma}) + \operatorname{Var}(\boldsymbol{\varepsilon}) \tag{4}$$

$$= \mathbf{Z}\operatorname{Var}(\boldsymbol{\gamma})\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n \tag{5}$$

$$= \sigma_\gamma^2 \mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n. \tag{6}$$

$\mathbf{V}$ can be given as a block diagonal,

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{V}_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{V}_b \end{bmatrix},$$

where

$$\mathbf{V}_i = \sigma_\varepsilon^2 \mathbf{I}_{k_i} + \sigma_\gamma^2 \mathbf{1}_{k_i} \mathbf{1}_{k_i}',$$

and

$$\mathbf{V}_i = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\varepsilon^2 + \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 & \sigma_\varepsilon^2 + \sigma_\gamma^2 \end{bmatrix}.$$

As a result, the variance-covariance matrix $\mathbf{V}_i$ of all observations within one block is compound symmetric: the main diagonal of the matrix contains the variances of the observations, while the off-diagonal elements are covariances. However, $\mathbf{V}_i$ can be rewritten as

$$\mathbf{V}_i = \sigma_\varepsilon^2 \left(\mathbf{I}_{k_i \times k_i} + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2} \mathbf{1}_{k_i} \mathbf{1}_{k_i}' \right) \tag{7}$$

$$= \sigma_\varepsilon^2 \left(\mathbf{I}_n + \eta \mathbf{Z}\mathbf{Z}' \right), \tag{8}$$

where $\eta = \sigma_\gamma^2 / \sigma_\varepsilon^2$ is a measure for the extent to which observations within the same block are correlated. The larger this variance ratio, the stronger observations within the same block are correlated.

When the random error terms as well as the group effects are normally distributed, the maximum likelihood estimate of the unknown model parameter $\boldsymbol{\beta}$ in Equation (1) is the generalised least squares (GLS) estimate. Detecting the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, requires to minimise

$$\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right)' \mathbf{V}^{-1} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right) = \mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} \tag{9}$$

with respect to $\boldsymbol{\beta}$, which is tantamount to detecting $\hat{\boldsymbol{\beta}}$, so that

$$\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right) \hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{10}$$

Therefore, the generalised least squares (GLS) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \tag{11}$$

and the variance-covariance matrix of the estimators is given by

$$\begin{aligned} \mathrm{Var}\left(\hat{\boldsymbol{\beta}} \right) &= \mathrm{Var}\left(\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \right) \right) \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \mathrm{Var}\left(\mathbf{Y} \right) \left(\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1} \right)' \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right) \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \right)^{-1}. \end{aligned} \tag{12}$$

Often, the variances σ_γ^2 and σ_ε^2 are not known and therefore, Equation (11) and Equation (12) cannot be used directly. Instead, the estimates of the variance components, $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\varepsilon^2$, are substituted in the GLS estimator as in Equation (11), yielding

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \tag{13}$$

where

$$\mathbf{V} = \hat{\sigma}_{\varepsilon}^{2}\mathbf{I}_{n} + \hat{\sigma}_{\gamma}^{2}\mathbf{Z}\mathbf{Z}'. \tag{14}$$

In that case, the variance-covariance matrix in Equation (12) can be approximated by

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}. \tag{15}$$

The generalised least square (GLS) estimator is unbiased, meaning that $\mathbb{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$, and is equal to the maximum likelihood estimator (MLE). The likelihood function defined as it is the joint probability density function for the observed data examined as a function of the parameters. Hence, the likelihood function for $\mathbf{Y}$ in Equation (1) is

$$L\left(\boldsymbol{\beta} \mid Y\right) = \left(2\pi\right)^{-n/2}\left|\mathbf{V}\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{V}^{-1}\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right], \tag{16}$$

where π is a constant which does not depend on $\boldsymbol{\beta}$. The maximum likelihood estimator (MLE) is the estimator that maximises the likelihood function, which is tantamount to detecting the $\hat{\boldsymbol{\beta}}$ as

$$\frac{\partial}{\partial\boldsymbol{\beta}}L\left(\hat{\boldsymbol{\beta}} \mid \mathbf{Y}\right) = 0, \tag{17}$$

which is equal to

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ln L\left(\hat{\boldsymbol{\beta}} \mid \mathbf{Y}\right) = 0, \tag{18}$$

where $\ln L\left(\hat{\boldsymbol{\beta}} \mid \mathbf{Y}\right)$ is the log likelihood function. As Equation (9) is proportionate to log of Equation (16), the GLS estimator in Equation (11) is the result of Equation (17) and Equation (18).

The restricted maximum likelihood (REML) used to estimate σ_{ε}^{2} and σ_{γ}^{2} is

$$l_{\mathrm{REML}}\left(\sigma_{\varepsilon}^{2},\sigma_{\gamma}^{2};\mathbf{Y}\right) = -\frac{1}{2}\ln\left|\mathbf{V}\right| - \frac{1}{2}\ln\left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right| - \frac{1}{2}\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\mathbf{V}^{-1}\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right), \tag{19}$$

where $\hat{\boldsymbol{\beta}}$ is defined in Equation (13). The restricted log-likelihood $l_{\mathrm{REML}}\left(\sigma_{\varepsilon}^{2},\sigma_{\gamma}^{2};\mathbf{Y}\right)$ is minimised with respect to the variance components σ_{ε}^{2} and σ_{γ}^{2} to obtain an unbiased estimate for the variance components.

3. The Bayesian Methodology

3.1. Bayesian Framework

The essential philosophy behind Bayesian inference is to update a prior distribution for an unidentified parameter to a posterior distribution by Bayes' theorem. Bayes' theorem can be used to estimate the conditional distributions. While the frequentist approach treats the parameters as unknown and fixed, the Bayesian approach regards them as random variables. We can define the prior distribution $p\left(\boldsymbol{\theta}\right)$ as the probability density (or mass) function which reflects our beliefs about $\boldsymbol{\theta}$ in the parameter space $\boldsymbol{\Theta}$. For given data $\mathbf{y} = \left(y_1, y_2, \cdots, y_n\right)'$,

the likelihood function $f(\mathbf{y}|\boldsymbol{\theta})$ can then be defined given the parameter $\boldsymbol{\theta}$ for the data $\mathbf{y}$. Also, we can define the posterior density (or mass) function $p(\boldsymbol{\theta}|y_1, y_2, \cdots, y_n)$ which represents our updated belief about $\boldsymbol{\theta}$ given the observed data $\mathbf{y}$.

Using Bayes' theorem, the posterior density of $\boldsymbol{\theta}$ given $\mathbf{y}$ is:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}. \tag{20}$$

Bayesian inference continues from this distribution. The denominator of Equation (20) is the marginal likelihood of $\mathbf{y}$, and it often does not need to be calculated because it is independent of $\boldsymbol{\theta}$. Bayes' rule can then be written as:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta}). \tag{21}$$

Equation (21) defines the unnormalised posterior density. The posterior then is proportional to the likelihood × the prior. For more details on Bayesian inference, see [12] and [13].

A prior distribution can be selected based on past information or experimental practice. It can be informative or uninformative. The informative distribution is given numerical information to estimate the parameter of concern. The uninformative reflects equilibrium among outcomes when weak information about the parameter is presented. There are two types of uninformative priors: proper prior and improper prior. The density for proper prior distribution integrates to 1 whereas the integral of the density for an improper distribution is not finite. If the prior integrates to any positive finite value, it is called an unnormalised density and can be renormalised-multiplied by a constant to integrate to 1 [12] [13].

3.2. Markov Chain Monte Carlo (MCMC) Methods

Markov Chain Monte Carlo simulation is a general method based on drawing values of the $\boldsymbol{\theta}$ from approximate distributions, and then correcting those draws to better approximate the target posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ [12] [13] [14]. A Markov chain can be defined as a sequence of random variables $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \cdots$ for which for any iteration t, the distribution of $\boldsymbol{\theta}^t$ depends only on the most recent value $\boldsymbol{\theta}^{t-1}$ [12] [13] [14]. A Markov chain is generated by sampling $\boldsymbol{\theta}^t \sim p(\boldsymbol{\theta}|\boldsymbol{\theta}^{t-1})$. This p is called the transition kernel of the Markov chain. Therefore, $\boldsymbol{\theta}^t$ depends only on $\boldsymbol{\theta}^{t-1}$, not on $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1, \cdots, \boldsymbol{\theta}^{t-2}$.

As $t \to \infty$, the sampling from Markov chain converges to the posterior for the right choice of transition kernel $p(\boldsymbol{\theta}|\mathbf{y})$. Thus, we should run the simulation long enough so that the distribution of the current draws is close enough to $p(\boldsymbol{\theta}|\mathbf{y})$.

3.2.1. Metropolis-Hastings Sampling

Metropolis-Hastings sampling was proposed by [15] and [16]. The theory of this sampling is based on rejection sampling. The acceptance-rejection method is a technique of getting samples from a distribution with an unknown form. The Metropolis-Hastings algorithm is a common expression for a family of Markov

chain simulation methods. It is worth describing the Metropolis algorithm first, then broadening it to discuss the Metropolis-Hastings algorithm. Let $p(\boldsymbol{\theta}|\mathbf{y})$ be the conditional posterior distribution where we want to sample from. Let $\boldsymbol{\theta}^t$ be the current parameter value, and let $\pi(\boldsymbol{\theta})$ be the proposal density. The proposal density is much like a conventional transition operator for a Markov chain, the proposal distribution depends only on the previous state in the chain. However, the transition operator for the Metropolis algorithm has a additional step that assesses whether or not the target distribution has a sufficiently large density near the proposed state to warrant accepting the chain.

3.2.2. Gibbs Sampling

A Gibbs sampler is the simplest of the Markov chain simulation algorithms, and it is used to sample from the conditional conjugate models, where we can directly sample from each conditional posterior [12] [13]. It is rare to find all the conditional posteriors in a model in known forms. One may find some conditional posterior distributions that are possible to directly sample from. Furthermore, one may find some of the conditional posteriors that cannot be straightforwardly sampled from. Therefore, the procedure for this issue is to update the parameters one at a time with the Gibbs sampler used where possible, and one-dimensional Metropolis updating where necessary. This process is called the Metropolis-Hastings within Gibbs sampling and will be used in this work.

4. A Hierarchical Mixture Model for Variable Selection

The linear mixed model fitted to data from a split-plot experiment with n responses is

$$\mathbf{y} \sim N\left(\beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\right), \tag{22}$$

where $\mathbf{y}$ is $n\times 1$ vector of random responses, β_0 is the intercept, $\mathbf{1}_n$ is a $n\times 1$ vector of ones, $\mathbf{X}$ is the $n\times p$ model matrix without the column of the intercept, $\boldsymbol{\beta}$ is the $p\times 1$ vector of fixed effect parameters and $\mathbf{V}$ is

$$\mathbf{V} = \sigma_\varepsilon^2\left(\mathbf{I}_n + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\mathbf{Z}\mathbf{Z}'\right),$$

where $\mathbf{Z}$ is the random effect design matrix. As $\rho = \frac{\sigma_\gamma^2}{\sigma^2}$, and $\sigma^2 = \sigma_\varepsilon^2 + \sigma_\gamma^2$, then $\mathbf{V}$ can be written as

$$\mathbf{V} = \sigma^2(1-\rho)\left(\mathbf{I}_n + \frac{\rho}{1-\rho}\mathbf{Z}\mathbf{Z}'\right). \tag{23}$$

We need to find the highest posterior probability of an indicator vector $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_p)$ such that

$$\nu_j = \begin{cases} 0 & \text{if } \beta_j = 0 \\ 1 & \text{if } \beta_j \neq 0 \end{cases}$$

for $j = 1, 2, \cdots, p$. When $\nu_j = 1$ the term is assumed to be active and will be in-

cluded in the model, and when $\nu_j = 0$ the term is assumed to be non-active and will not be included in the model.

Following [6], and [5], we assume that $\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\nu}, c \sim \mathbf{N}\left(\mathbf{0}_p, \sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)$, where $\boldsymbol{\nu}$ is the indicator vector, c is the prior variance of the slab distribution, and $\mathbf{D}_{\boldsymbol{\nu},c}$ is a diagonal matrix with the j^{th} diagonal element $cI\left(\nu_j = 1\right) + dI\left(\nu_j = 0\right), j = 1, \cdots, p$. The parameters $\sigma^2, \boldsymbol{\nu}$ and c will be given prior distributions, and the parameter d is assumed to be a small fixed non-negative number because we want the spike distribution to have a smaller variance than the slab distribution. Formally the prior construction of $\boldsymbol{\beta}$ is the following:

$$\beta_j \mid \sigma^2, \nu_j, c \sim \left(1-\nu_j\right)\mathbf{N}\left(0, d\sigma^2\right) + \nu_j \mathbf{N}\left(0, c\sigma^2\right).$$

For every coefficient β_j, a Bernoulli variable ν_j is defined taking values 1 and 0 with probability of inclusion ω, as $p\left(\nu_j = 1\right) = \omega$ and $p\left(\nu_j = 0\right) = \left(1-\omega\right)$. Often, ν_j's are taken as independent Bernoulli (ω) random variables, where $0 < \omega < 1$. It is common to fix ω in the normal mixture, however, we shall deal with ω as a parameter to investigate different values of ω, and sample it from the Beta distribution as it will be explained in Section 6.3.

5. Prior Distributions

1) Following the prior distributions used by [5], we assume that the prior distribution for the fixed effects is $\boldsymbol{\beta} \sim \mathbf{N}\left(0, \sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)$

$$p\left(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\nu}, c\right) \propto \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1}\boldsymbol{\beta}\right).$$

2) The prior distribution for the total variance is $\sigma^2 \sim \mathbf{IG}\left(a, b\right)$,

$$p\left(\sigma^2\right) \propto \left(\sigma^2\right)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right).$$

For this work, we used $a = 0$ and $b = 0$ following [5] as this yields the common non-informative prior for σ^2. This prior is improper, however we will sample from the posterior distribution, which should be a proper gamma distribution.

3) The prior distribution for the correlation parameter is $\rho \sim \mathbf{Beta}\left(a', b'\right)$ with shape parameters $a', b' > 0$. We consider $a' = b' = 2$ following [8]. According to [8], "A $\mathbf{Beta}\left(a', b'\right)$ prior distribution for a correlation parameter can be interpreted as indicating a prior point estimate of $a'/\left(a'+b'\right)$, this prior information being worth $a'+b'$ observations". Our prior was selected to be centred at $2/\left(2+2\right) = 0.5$ and to be worth four observations in each block. For an experiment with $a'+b'$ observations, the posterior distribution would give equal weight to the prior and the likelihood [8]. Similar choice of the prior is in [6] as they set $a = b = 2$ so that the $p\left(\rho\right)$ is symmetric with a mode of 0.5. This has the effect of pulling the posterior mode of $p\left(\rho\right)$ towards 0.5 when the data are scarce. The prior density for ρ is

$$p\left(\rho\right) \propto \rho^{\left(a'-1\right)}\left(1-\rho\right)^{\left(b'-1\right)}.$$

4) The prior distribution for the elements of the indicator vector is

$\nu_j \sim \textbf{Bernoulli}(\omega)$,

$$p(\nu_j) = \begin{cases} \omega & \text{if } \nu_j = 1 \\ 1-\omega & \text{if } \nu_j = 0 \end{cases}$$

where ω is the prior probability that β_j is active following [5].

5) The prior distribution for the elements of the probability of inclusion is $\omega \sim \textbf{Beta}(c_0, d_0)$,

$$p(\omega) \propto \omega^{(c_0-1)}(1-\omega)^{(d_0-1)}.$$

[5] set $\omega = 0.25$. However, we select c_0 and d_0 such that the prior of ω has a mode = 0.25. The choice of $c_0 = 2$ and $d_0 = 4$ results in a prior with a mode = 0.25 and the upper cumulative percentile at 5% equals 0.66. Meaning a 5% chance the observations have a probability density function ≥ 0.66.

6) The prior distribution for the slab variance c is a discrete uniform prior distribution with support points T = {1/4, 9/16, 1, 4, 9, 16, 25} as given by [5]. They found that large values of c tend to favor sparse models with large effects and in this case small effects will be missed. On the other hand, small values of c tend to favor less sparse models. Moreover, very small values of c tend to favor sparse models again. They select the support points in T such that it covers small and large values of c. The prior distribution for c is

$$p(c) = \begin{cases} \frac{1}{7} & \text{if } c \in \mathrm{T} \\ 0 & \text{otherwise} \end{cases}$$

6. Full Conditional Distributions

We use the prior distributions presented in Section 5 to derive the full conditional distributions. The likelihood of the data depends on $\boldsymbol{\beta}$, so we can derive the conditional distribution for $\boldsymbol{\beta}$ using the prior distribution $\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\nu}, c \sim \mathbf{N}(\mathbf{0}_p, \sigma^2 \mathbf{D}_{\boldsymbol{\nu},c})$ and the likelihood for the model 22

$$L(y \mid \boldsymbol{\theta}) \propto |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right].$$

Note that we standardise both $\mathbf{X}$ and $\mathbf{y}$ so the fixed effect vector $\boldsymbol{\beta}$ does not include the intercept. The conditional distribution for $\boldsymbol{\beta}$ can be expressed as:

$$\begin{aligned} p(\boldsymbol{\beta} \mid \boldsymbol{\nu}, \sigma^2, c, \omega, \rho, \mathbf{y}) &\propto p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, c, \omega, \rho)\, p(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\nu}, c) \\ &\propto |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] \\ &\quad \times |\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}'(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c})^{-1}\boldsymbol{\beta}\right) \\ &\propto |\mathbf{V}|^{-1/2} |\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}|^{-1/2} \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}'(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c})^{-1}\boldsymbol{\beta}\right)\right. \\ &\quad \left. -\frac{1}{2}\left(\mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}\right)\right] \end{aligned}$$

$$\begin{aligned}
&\propto |\mathbf{V}|^{-1/2} \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta}\right] \\
&\times \exp\left[-\frac{1}{2}\left(-\mathbf{y}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta}\right] \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right] \\
&\propto |\mathbf{V}|^{-1/2} \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta}\right] \\
&\times \exp\left[\frac{1}{2}\left(\mathbf{y}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right] \\
&\propto |\mathbf{V}|^{-1/2} \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta} + \boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right].
\end{aligned}$$

The key to deriving the joint posterior distribution is to rewrite the expression in the exponential part in a more convenient form. This can happen by using the multivariate completion of squares:

$$\mathrm{U}'\mathrm{AU} - 2\mathrm{U}'\boldsymbol{\alpha} = \left(\mathrm{U} - \mathrm{A}^{-1}\boldsymbol{\alpha}\right)' \mathrm{A}\left(\mathrm{U} - \mathrm{A}^{-1}\boldsymbol{\alpha}\right) - \boldsymbol{\alpha}'\mathrm{A}^{-1}\boldsymbol{\alpha},$$

where A is a symmetric positive definite (hence invertible) matrix. We assume $\mathrm{U} = \boldsymbol{\beta}$, $\mathrm{A} = \left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$, and $\alpha = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$.

The conditional distribution for $\boldsymbol{\beta}$ can be written as:

$$\begin{aligned}
&p\left(\boldsymbol{\beta} \mid \boldsymbol{\nu}, \sigma^2, c, \omega, \rho, \mathbf{y}\right) \\
&\propto |\mathbf{V}|^{-1/2} \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right] \\
&\propto |\mathbf{V}|^{-1/2} \left|\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right|^{-1/2} \exp\left[\left[\boldsymbol{\beta} - \left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right]'\right. \\
&\left.\times\left[\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right]\left[\boldsymbol{\beta} - \left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)\right]\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_*\right)' \mathrm{D}_*^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}_*\right)\right].
\end{aligned}$$

Thus, we can sample $\boldsymbol{\beta}$ from the conditional posterior $\mathrm{N}\left(\boldsymbol{\beta}_*, \mathrm{D}_*\right)$, where

$$\boldsymbol{\beta}_* = \left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right) \text{ and } \mathrm{D}_* = \left(\left(\sigma^2 \mathbf{D}_{\boldsymbol{\nu},c}\right)^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}. \quad (24)$$

6.1. The Conditional Distribution for ρ

The likelihood of the data depends on ρ, so the conditional distribution for ρ can be derived by

$$\begin{aligned}
&p\left(\rho \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, c, \omega, \mathbf{y}\right) \propto p\left(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, c, \omega, \rho\right) p\left(\rho\right) \\
&\propto |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right] \times \rho^{(a'-1)}\left(1-\rho\right)^{(b'-1)}.
\end{aligned}$$

We note here that the likelihood depends on ρ through $\mathbf{V}$ as in (4.6), so we can express $\mathbf{V}$ as a function of ρ,

$$\mathrm{V} = \sigma^2\left(1-\rho\right)\left(\mathbf{I}_n + \frac{\rho}{1-\rho}\mathbf{Z}\mathbf{Z}'\right).$$

The conditional distribution for ρ is a non-standard distribution that cannot be sampled directly. Therefore, we use the Metropolis-Hastings (M-H) rejection sampling. Our correlation parameter is $\rho \in (0,1]$, and has a prior $\beta(a',b')$. We apply the Random-Walk Metropolis-Hastings algorithm, and select a proposal distribution of log-normal distribution for the variance ratio η where $\eta = f(\rho) = \frac{\rho}{1-\rho}$ with a mean equal to the current value of η^t at iteration t and variance s^2. The choice of s^2 affects the jumping rule in the random walk proposal distribution. As we have one parameter to be updated in the random walk algorithm which is ρ, we follow [12] and [17] to set $s^2 = g^2\Sigma$. The most efficient jump has a scale $g \approx 2.4/\sqrt{h}$ where h is the number of parameters which will be updated. In this work, we set $g = 2.4$ and $h = 1$ following [17], and we set $\Sigma = 100$ as this yields an appropriate acceptance rate associated with the independent sampler of the ACF plot. Thus, $\eta \in (0,\infty)$ and

$$g(\eta) = \frac{1}{\eta\sqrt{2\pi s^2}} \exp\left[-\frac{1}{2s^2}\left(\ln\eta - \eta^t\right)^2\right]$$

We can use $\eta = \frac{\rho}{1-\rho}$ as a transformation function between η and ρ as $\rho = \frac{\eta}{1+\eta}$ and the Jacobian function of ρ is $J(\rho) = \frac{\mathrm{d}\eta}{\mathrm{d}\rho} = \frac{1}{(1-\rho)^2}$.

We draw a proposal value η_* from a log-normal(η^t, s^2) distribution, and the probability of accepting or rejecting η_* is the minimum of 1 and the ratio r where r is

$$r = \frac{p(\rho_* \mid \text{all})}{p(\rho^t \mid \text{all})} \times \frac{q(\eta^t \mid \eta_*)}{q(\eta_* \mid \eta^t)},$$

which is equivalent to

$$r = \frac{p(\rho_* \mid \text{all})}{p(\rho^t \mid \text{all})} \times \frac{q(\rho^t \mid \rho_*) J(\rho^t)}{q(\rho_* \mid \rho^t) J(\rho_*)}.$$

Our proposal ratio is

$$\frac{q(\eta^t \mid \eta_*)}{q(\eta_* \mid \eta^t)} = \frac{\eta_* \exp\left[-\frac{1}{2s^2}\left(\ln\eta^t - \eta_*\right)^2\right]}{\eta^t \exp\left[-\frac{1}{2s^2}\left(\ln\eta_* - \eta^t\right)^2\right]},$$

which is equivalent to

$$\frac{q(\rho^t \mid \rho_*) J(\rho^t)}{q(\rho_* \mid \rho^t) J(\rho_*)} = \frac{\left(\frac{\rho_*}{1-\rho_*}\right) \exp\left[-\frac{1}{s^2}\left(\ln\left(\frac{\rho^t}{1-\rho^t}\right) - \left(\frac{\rho_*}{1-\rho_*}\right)\right)^2\right] \times \left|\frac{1}{(1-\rho^t)^2}\right|}{\left(\frac{\rho^t}{1-\rho^t}\right) \exp\left[-\frac{1}{s^2}\left(\ln\left(\frac{\rho_*}{1-\rho_*}\right) - \left(\frac{\rho^t}{1-\rho^t}\right)\right)^2\right] \times \left|\frac{1}{(1-\rho_*)^2}\right|}.$$

The ratio r can be expressed as

$$r=\frac{\left|\mathbf{V}\left(\rho_*\right)\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{V}\left(\rho_*\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\times\left(\rho_*\right)^{(a'-1)}\left(1-\rho_*\right)^{(b'-1)}}{\left|\mathbf{V}\left(\rho^t\right)\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{V}\left(\rho^t\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\times\left(\rho^t\right)^{(a'-1)}\left(1-\rho^t\right)^{(b'-1)}}$$
$$\times\frac{\rho_*\left(1-\rho^t\right)\exp\left[-\frac{1}{s^2}\left(\ln\left(\frac{\rho^t}{1-\rho^t}\right)-\left(\frac{\rho_*}{1-\rho_*}\right)\right)^2\right]\left|\left(1-\rho^t\right)^{-2}\right|}{\rho^t\left(1-\rho_*\right)\exp\left[-\frac{1}{s^2}\left(\ln\left(\frac{\rho_*}{1-\rho_*}\right)-\left(\frac{\rho^t}{1-\rho^t}\right)\right)^2\right]\left|\left(1-\rho_*\right)^{-2}\right|}. \quad (25)$$

where $\mathrm{V}\left(\rho_*\right)=\sigma^2\left(1-\rho_*\right)\left(\mathbf{I}_n+\frac{\rho_*}{1-\rho_*}\mathbf{ZZ}'\right)$, and

$\mathrm{V}\left(\rho^t\right)=\sigma^2\left(1-\rho^t\right)\left(\mathbf{I}_n+\frac{\rho^t}{1-\rho^t}\mathbf{ZZ}'\right)$.

6.2. The Conditional Distribution for σ^2

The likelihood of the data depends on σ^2, so we can express the conditional distribution of σ^2 as

$$p\left(\sigma^2\mid\boldsymbol{\beta},\rho,\boldsymbol{\nu},c,\omega,\mathbf{y}\right)\propto p\left(\mathbf{y}\mid\boldsymbol{\beta},\boldsymbol{\nu},\sigma^2,c,\omega,\rho\right)p\left(\sigma^2\right)$$
$$\propto\left|\mathbf{V}\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\times\left(\sigma^2\right)^{-a-1}\exp\left(-\frac{b}{\sigma^2}\right).$$

We know that $\mathrm{V}=\sigma^2\left(1-\rho\right)\left(\mathbf{I}_n+\frac{\rho}{1-\rho}\mathbf{ZZ}'\right)$, so the conditional posterior for σ^2 can be written as

$$p\left(\sigma^2\mid\mathbf{y},\cdots\right)$$
$$\propto\left|\left(1-\rho\right)\sigma^2\left(\mathbf{I}_n+\frac{\rho}{1-\rho}\mathbf{ZZ}'\right)\right|^{-1/2}\left(\sigma^2\right)^{-a-1}\times\exp\left(-\frac{b}{\sigma^2}\right)$$
$$\times\exp\left[-\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\left(1-\rho\right)\sigma^2\left(\mathbf{I}_n+\frac{\rho}{1-\rho}\mathbf{ZZ}'\right)\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]$$
$$\propto\left(\sigma^2\right)^{-\left(a+\frac{n}{2}\right)-1}\exp\left(-\frac{1}{\sigma^2}\left[\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\left(1-\rho\right)\left(\mathbf{I}_n+\frac{\rho}{1-\rho}\mathbf{ZZ}'\right)\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{2}+b\right]\right).$$

This is the inverse gamma distribution with a shape parameter a^* and a scale parameter b^* such that

$$a^*=a+\frac{n}{2}\ \text{ and }\ b^*=\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\left(1-\rho\right)\left(\mathbf{I}_n+\frac{\rho}{1-\rho}\mathbf{ZZ}'\right)\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{2}+b. \quad (26)$$

The indicator vector can be drawn conditionally on the regressor coefficient and computation of the marginal likelihood is not required. The prior probabili-

ties for ν_j are

$$p(\nu_j)=\begin{cases}\omega & \text{if } \nu_j=1\\ 1-\omega & \text{if } \nu_j=0\end{cases}$$

where ω is the prior probability that β_j is active. The joint conditional posterior distribution for $\boldsymbol{\nu}$ has mass function

$$\begin{aligned}p(\boldsymbol{\nu}\mid\boldsymbol{\beta},y)&\propto p(\mathbf{y}\mid\boldsymbol{\beta},\boldsymbol{\nu})\,p(\boldsymbol{\beta},\boldsymbol{\nu})\\&=p(\mathbf{y}\mid\boldsymbol{\beta})\,p(\boldsymbol{\beta},\boldsymbol{\nu})\\&\propto p(\boldsymbol{\beta},\boldsymbol{\nu})\\&=p(\boldsymbol{\beta}\mid\boldsymbol{\nu})\,p(\boldsymbol{\nu}).\end{aligned}$$

The conditional density for $\boldsymbol{\beta}$ given $\boldsymbol{\nu}$ is

$$\begin{aligned}p(\boldsymbol{\beta}\mid\boldsymbol{\nu})\propto&\left|\sigma^2\operatorname{diag}\left[c\mathrm{I}(\nu_j=1)+d\mathrm{I}(\nu_j=0)\right]\right|^{-1/2}\\&\times\exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\sigma^2\operatorname{diag}\left[c\mathrm{I}(\nu_j=1)+d\mathrm{I}(\nu_j=0)\right]\right)^{-1}\boldsymbol{\beta}\right].\end{aligned}$$

The conditional distribution for the j^{th} component given ν_j is

$$\begin{aligned}p(\beta_j\mid\nu_j)\propto&\left|\sigma^2\left[c\mathrm{I}(\nu_j=1)+d\mathrm{I}(\nu_j=0)\right]\right|^{-1/2}\\&\times\exp\left[-\frac{\beta_j^2}{2\sigma^2\left[c\mathrm{I}(\nu_j=1)+d\mathrm{I}(\nu_j=0)\right]}\right].\end{aligned}$$

The conditional posterior probabilities for ν_j are therefore

$$\begin{aligned}p(\nu_j=1\mid\beta_j,\mathbf{y})&=p(\nu_j=1)\,p(\beta_j\mid\nu_j=1)\\&\propto\omega\left|c\sigma^2\right|^{-1/2}\exp\left[-\frac{\beta_j^2}{2c\sigma^2}\right],\end{aligned}\tag{27}$$

and

$$\begin{aligned}p(\nu_j=0\mid\beta_j,\mathbf{y})&=p(\nu_j=0)\,p(\beta_j\mid\nu_j=0)\\&\propto(1-\omega)\left|d\sigma^2\right|^{-1/2}\exp\left[-\frac{\beta_j^2}{2d\sigma^2}\right].\end{aligned}\tag{eq:4.10}$$

6.3. The Conditional Distribution for ω

The probability of inclusion ω can be drawn conditionally on the indicator and computation of the marginal likelihood is not required. Hence the conditional distribution for ω is

$$\begin{aligned}p(\omega\mid\boldsymbol{\nu},\sigma^2,c,\boldsymbol{\beta},\rho,\mathbf{y})&\propto p(\mathbf{y}\mid\omega,\boldsymbol{\nu},\sigma^2,c,\boldsymbol{\beta},\rho)\,p(\omega,\boldsymbol{\nu},\sigma^2,c,\boldsymbol{\beta},\rho)\\&=p(\mathbf{y}\mid\boldsymbol{\nu},\sigma^2,c,\boldsymbol{\beta},\rho)\,p(\omega,\boldsymbol{\nu})\\&\propto p(\omega,\boldsymbol{\nu})\\&=p(\boldsymbol{\nu}\mid\omega)\,p(\omega)\\&\propto\omega^{\sum_{j=1}^{p}\nu_j}(1-\omega)^{p-\sum_{j=1}^{p}\nu_j}\times\omega^{(c_0-1)}(1-\omega)^{(d_0-1)}\\&\propto\omega^{c_0+\sum_{j=1}^{p}\nu_j-1}(1-\omega)^{p-\sum_{j=1}^{p}\nu_j+d_0-1}.\end{aligned}$$

Hence,

$$\omega \mid \boldsymbol{\nu} \sim \text{Beta}\left(c_0 + \sum_{j=1}^{p} \nu_j, p - \sum_{j=1}^{p} \nu_j + d_0\right), \text{ where } p - \sum_{j=1}^{p} \nu_j = \sum_{j=1}^{p} \text{I}\left(\nu_j = 0\right). \quad (28)$$

6.4. The Conditional Distribution for *c*

The prior distribution for c is a discrete uniform distribution with support points T = {1/4, 9/16, 1, 4, 9, 16, 25}, and it can be drawn conditionally on the regressor coefficient. The computation of the marginal likelihood is not required. Hence, the conditional distribution for c

$$\begin{aligned}
p\left(c \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, \omega, \rho, \mathbf{y}\right) &\propto p\left(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, c\right) p\left(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, c\right) \\
&= p\left(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2\right) p\left(\boldsymbol{\beta}, \boldsymbol{\nu}, \sigma^2, c\right) \\
&\propto p\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu}, c\right) \\
&= p\left(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\nu}, c\right) p\left(c\right) \\
&\propto \frac{1}{7}\left|\sigma^2 \text{diag}\left[c\text{I}\left(\nu_j = 1\right) + d\text{I}\left(\nu_j = 0\right)\right]\right|^{-1/2} \\
&\quad \times \exp\left[-\frac{1}{2}\boldsymbol{\beta}'\left(\sigma^2 \text{diag}\left[c\text{I}\left(\nu_j = 1\right) + d\text{I}\left(\nu_j = 0\right)\right]\right)^{-1}\boldsymbol{\beta}\right] \\
&\propto \frac{1}{7}\left[\prod_{j=1}^{p}\left[c\text{I}\left(\nu_j = 1\right) + d\text{I}\left(\nu_j = 0\right)\right]\right]^{-1/2} \\
&\quad \times \exp\left[-\frac{1}{2c}\left(\boldsymbol{\beta}'\text{I}_{\sum_j \nu_j = 1}\boldsymbol{\beta}\right)\right] \times \exp\left[-\frac{1}{2d}\left(\boldsymbol{\beta}'\text{I}_{\sum_j \nu_j = 0}\boldsymbol{\beta}\right)\right] \\
&\propto \frac{1}{7}c^{\frac{-\sum_{j=1}^{p}\nu_j}{2}} + \frac{1}{7}d^{\frac{-\sum_{j=1}^{p}\left(1-\nu_j\right)}{2}} \times \exp\left[-\frac{1}{2c}\left(\boldsymbol{\beta}'\text{I}_{\sum_j \nu_j = 1}\boldsymbol{\beta}\right)\right] \\
&\quad \times \exp\left[-\frac{1}{2d}\left(\boldsymbol{\beta}'\text{I}_{\sum_j \nu_j = 0}\boldsymbol{\beta}\right)\right] \\
&\propto \frac{1}{7}c^{\frac{-\sum_{j=1}^{p}\nu_j}{2}} \times \exp\left[-\frac{1}{2c}\left(\boldsymbol{\beta}'\text{I}_{\sum_j \nu_j = 1}\boldsymbol{\beta}\right)\right] \\
&\propto \frac{1}{7}c^{\frac{-\sum_{j=1}^{p}\nu_j}{2}} \times \exp\left[-\frac{1}{2c}\left(\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right].
\end{aligned}$$

Therefore,

$$p\left(c \mid \mathbf{y}, \cdots\right) \propto \begin{cases} \frac{1}{7}c^{\frac{-\sum_{j=1}^{p}\nu_j}{2}} \times \exp\left[-\frac{1}{2c}\left(\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right] & \text{if } c \in \text{T} \\ 0 & \text{otherwise} \end{cases}$$

Then, the posterior probabilities of the conditional distribution $p\left(c \mid \mathbf{y}, \cdots\right)$ are

$$p\left(c = \frac{1}{4} \mid \mathbf{y}, \cdots\right) \propto \frac{1}{7}\left(\frac{1}{4}\right)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}} \times \exp\left[-\frac{1}{2\left(\frac{1}{4}\right)}\left(\boldsymbol{\beta}'\boldsymbol{\beta}\right)\right],$$

$$p\left(c=\frac{9}{16}\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}\left(\frac{9}{16}\right)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2\left(\frac{9}{16}\right)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right],$$

$$p\left(c=1\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}(1)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2(1)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right],$$

$$p\left(c=4\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}(4)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2(4)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right],$$

$$p\left(c=9\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}(9)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2(9)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right],$$

$$p\left(c=16\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}(16)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2(16)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right],$$

and

$$p\left(c=25\,|\,\mathbf{y},\cdots\right)\propto\frac{1}{7}(25)^{\frac{-\sum_{j=1}^{p}\nu_j}{2}}\times\exp\left[-\frac{1}{2(25)}(\boldsymbol{\beta}'\boldsymbol{\beta})\right].$$

The conditional posterior $p\left(c\,|\,\mathbf{y},\cdots\right)$ can be written as

$$p\left(c\,|\,\mathbf{y},\cdots\right)=\begin{cases}\dfrac{c^{-\frac{\sum_{j=1}^{p}\nu_j}{2}}\exp\left[-\dfrac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2c}\right]}{\sum_{c\in T}c^{-\frac{\sum_{j=1}^{p}\nu_j}{2}}\exp\left[-\dfrac{\boldsymbol{\beta}'\boldsymbol{\beta}}{2c}\right]} & \text{if } c\in \mathrm{T}\\ 0 & \text{otherwise}\end{cases} \tag{29}$$

7. Bayesian Variable Selection Algorithms

In this section, we introduce the computational algorithms which we use in the Bayesian analysis for variable selection using the SSVS, and the SSVS-SPD. In this work, we choose an asymmetric proposal distribution, the log-normal density. We apply the Metropolis-Hastings algorithms to sample the variance ratio η where $\eta=f(\rho)=\frac{\rho}{1-\rho}$, and ρ is the correlation parameter. This is because of the fact that in our experiments, observations from different subplots within the same wholeplot are positively correlated as $\rho=\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2+\sigma_\gamma^2}$; also observations from different wholeplots are independent [5].

7.1. The Stochastic Search Variable Selection (SSVS) Algorithm

We process the MCMC estimation of the parameters $\boldsymbol{\beta},\rho,\sigma^2,\boldsymbol{\nu},\omega$ and c. We use the priors of all these parameters as in Section 5. The following Metropolis-Hastings within Gibbs sampling algorithm can be implemented. Let $\mathbf{y}$ be the $n\times 1$ vector of random responses, $\mathbf{X}$ is the $n\times p$ model matrix without the

column of the intercept, $\boldsymbol{\beta}$ is the $p\times 1$ vector of fixed effect parameters, where p is the number of fixed effect parameters that need to be estimated. We set initial values for the parameters as $\boldsymbol{\beta}^{(0)}=\mathbf{1}_p$, $\boldsymbol{\nu}^{(0)}=\mathbf{1}_p$, $\rho^{(0)}=0.5$, $\sigma^{2(0)}=10$, $c^{(0)}=1$, $\omega^{(0)}=0.5$, $d=0.001$. Starting at the tth iteration such that $t=1,2,\cdots,its$ where $its=10000$, and setting $j=1,2,\cdots,p$, the sampling algorithm is:

1) For $j=1,2,\cdots,p$, sample $\nu_j^{(t)}$ of the indicator vector $\boldsymbol{\nu}^{(t)}$ using (27) for $\beta_j^{(t-1)}$, $c^{(t-1)}$, $\sigma^{2(t-1)}$, and $\omega^{(t-1)}$.

2) Sample the mixture weight $\omega^{(t)}$ using (28) for $\boldsymbol{\nu}^{(t)}$.

3) Sample the regressor coefficients $\boldsymbol{\beta}^{(t)}$ using (24) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{V}^{(t-1)}$, $\mathbf{D}^{(t-1)}$, $c^{(t-1)}$, $\boldsymbol{\nu}^{(t)}$, and $\sigma^{2(t-1)}$.

4) Sample the total variance $\sigma^{2(t)}$ using (26) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{Z}$, $\boldsymbol{\beta}^{(t)}$, and $\rho^{(t-1)}$.

5) a) Sample $\rho_*^{(t)}$ from $\beta(a',b')$;

b) Calculate $\alpha^{(t)}$ using (25) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{V}\left(\rho_*^{(t)}\right)$, $\mathbf{V}\left(\rho^{(t-1)}\right)$, and $\boldsymbol{\beta}^{(t)}$;

c) Sample $u^{(t)}$ from $U(0,1)$;

d) If $\alpha^{(t)}>u^{(t)}$, then set $\rho^{(t)}=\rho_*^{(t)}$, otherwise set $\rho^{(t)}=\rho^{(t-1)}$.

6) Sample $c^{(t)}$ from the set T with probability mass function given in (29) for $\boldsymbol{\beta}^{(t)}$, and $\boldsymbol{\nu}^{(t)}$.

7.2. The Stochastic Search Variable Selection for Split-Plot Design (SSVS-SPD) Algorithm

We adapt the SSVS for the analysis of data from split-plot designs by taking into account the two types of factors, *i.e.* the whole-plot factors and the subplot factors which expected to have different effect sizes for the two strata in split-plot design [9]. This approach can be reported as the Stochastic Search Variable Selection for Split-Plot Design (SSVS-SPD). While the SSVS samples all parameters from one slab variance posterior distribution, the SSVS-SPD samples the whole-plot parameters and the subplot parameters from two different slab variance posterior distributions given that the whole-plot and the subplot effects might have different sizes. While the SSVS samples all parameters from one slab variance posterior distribution, the SSVS-SPD samples the whole-plot parameters and the subplot parameters from two different slab variance posterior distributions. We use the same priors as in the SSVS for all the parameters of interest as in Section 5. Basically, the SSVS-SPD can be seen as running the SSVS twice in one process, one for subplot factors and the other one for whole-plot factors. The algorithm can be explained as follows:

We process the MCMC estimation of the parameters $\boldsymbol{\beta},\rho,\sigma^2,\boldsymbol{\nu},\omega$ and c. The following Metropolis-Hastings within Gibbs sampling algorithm can be implemented. Let $\mathbf{y}$ be the $n\times 1$ vector of random responses, $\mathbf{X}$ is the $n\times p$ model matrix without the column of the intercept, $\mathbf{X.S}$ is the $n\times p_s$ model matrix for subplot factors where p_s is the number of subplot fixed effect parameters.

Also, $\mathbf{X.W}$ is the $n\times p_w$ model matrix for whole-plot factors where p_w is the number of whole-plot fixed effect parameters. The $\boldsymbol{\beta}=(\boldsymbol{\beta}_s,\boldsymbol{\beta}_w)$ is the

$p \times 1$ vector of fixed effect parameters, where p is the number of fixed effect parameters that need to be estimated, $\boldsymbol{\beta}_s$ is the $p_s \times 1$ subplot effect parameters, and $\boldsymbol{\beta}_w$ is the $p_w \times 1$ whole-plot effect parameters.

We set initial values for the parameters as $\boldsymbol{\beta}_s^{(0)} = \mathbf{1}_{p_s}$, $\boldsymbol{\beta}_w^{(0)} = \mathbf{1}_{p_w}$. The initial values for the indicator vectors for the subplot factor $\boldsymbol{\nu}_s$ and the whole-plot factor $\boldsymbol{\nu}_w$ are $\boldsymbol{\nu}_s^{(0)} = \mathbf{1}_{ps}$ and $\boldsymbol{\nu}_w^{(0)} = \mathbf{1}_{pw}$. Also, $\rho^{(0)} = 0.5$, $\sigma^{2(0)} = 10$, and $d = 0.001$. The initial values for the slab variance for the subplot factors c_s and for the slab variance for the whole-plot factors c_w are $c_s^{(0)} = c_w^{(0)} = 1$. Finally, the initial weights for the subplot factors ω_s and for the whole-plot factors ω_w are $\omega_s^{(0)} = \omega_w^{(0)} = 0.5$.

Starting at the tth iteration such that $t = 1, 2, \cdots, its$ where $its = 10000$, and setting $j = 1, 2, \cdots, p_s$ and $k = 1, 2, \cdots, p_w$, the sampling algorithm is:

1) For $j = 1, 2, \cdots, p_s$, and $k = 1, 2, \cdots, p_w$ sample $\nu_{sj}^{(t)}$ and $\nu_{wk}^{(t)}$ of the indicator vectors $\boldsymbol{\nu}_s^{(t)}$ and $\boldsymbol{\nu}_w^{(t)}$ using (27) for $\beta_{sj}^{(t-1)}$, $\beta_{wk}^{(t-1)}$, $c_s^{(t-1)}$, $c_w^{(t-1)}$, $\sigma^{2(t-1)}$, $\omega_s^{(t-1)}$, and $\omega_w^{(t-1)}$.

2) Allocate $\boldsymbol{\nu}^{(t)} = \left(\boldsymbol{\nu}_s^{(t)}, \boldsymbol{\nu}_w^{(t)}\right)$.

3) Sample the mixture weights $\omega_s^{(t)}$ and $\omega_w^{(t)}$ using (28) for $\boldsymbol{\nu}_s^{(t)}$ and $\boldsymbol{\nu}_w^{(t)}$.

4) Allocate $\omega^{(t)} = \left(\omega_s^{(t)}, \omega_w^{(t)}\right)$.

5) Sample the regressor coefficients $\boldsymbol{\beta}_s^{(t)}$ and $\boldsymbol{\beta}_w^{(t)}$ using (24) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{V}^{(t-1)}$, $\mathbf{D}_s^{(t-1)}$, $\mathbf{D}_w^{(t-1)}$, $c_s^{(t-1)}$, $c_w^{(t-1)}$, $\boldsymbol{\nu}_s^{(t)}$, $\boldsymbol{\nu}_w^{(t)}$, and $\sigma^{2(t-1)}$. Where the $\mathbf{D}_s$ is a diagonal matrix with the j^{th} diagonal element $c_s^{(t-1)} I\left(\nu_{sj} = 1\right) + dI\left(\nu_{sj} = 0\right)$, and $\mathbf{D}_w$ is a diagonal matrix with the k^{th} diagonal element $c_w^{(t-1)} I\left(\nu_{wk} = 1\right) + dI\left(\nu_{wk} = 0\right)$.

6) Allocate $\boldsymbol{\nu}^{(t)} = \left(\boldsymbol{\nu}_s^{(t)}, \boldsymbol{\nu}_w^{(t)}\right)$ and $\mathbf{D}^{(t)} = \text{diag}\left(\mathbf{D}_s^{(t)}, \mathbf{D}_w^{(t)}\right)$.

7) Sample the total variance $\sigma^{2(t)}$ using (26) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{Z}$, $\boldsymbol{\beta}^{(t)}$, and $\rho^{(t-1)}$.

8) a) Sample $\rho_*^{(t)}$ from $\beta(a, b)$;

b) Calculate $\alpha^{(t)}$ using (25) for $\mathbf{X}$, $\mathbf{y}$, $\mathbf{V}\left(\rho_*^{(t)}\right)$, $\mathbf{V}\left(\rho^{(t-1)}\right)$, and $\boldsymbol{\beta}^{(t)}$;

c) Sample $u^{(t)}$ from $U(0,1)$;

d) If $\alpha^{(t)} > u^{(t)}$, then set $\rho^{(t)} = \rho_*^{(t)}$, otherwise set $\rho^{(t)} = \rho^{(t-1)}$.

9) Sample $c_s^{(t)}$ and $c_w^{(t)}$ from the set T with probability mass function given in (29) for $\boldsymbol{\beta}_s^{(t)}$, $\boldsymbol{\beta}_w^{(t)}$, $\boldsymbol{\nu}_s^{(t)}$ and $\boldsymbol{\nu}_w^{(t)}$.

10) Allocate $c^{(t)} = \left(c_s^{(t)}, c_w^{(t)}\right)$.

8. Real Data Application

[18] and [19] described an experiment in the production of vinyl for automobile seat covers. The experiment has 28 runs and is a modified version of an example in [19]. It involves the production of vinyl for automobile seat covers. In the experiment, the effects of five factors on the thickness of the vinyl are investigated. Three of the factors are mixture components and two of them are so-called process variables. As in ordinary mixture experiments, the component proportions sum to one. In this example, the response of interest does not only depend on these proportions, but also on the effects of the process variables. The mixture components in the experiment are three plasticizers whose proportions are

represented by s_1, s_2 and s_3. The two process variables studied are rate of extrusion (w_1) and temperature of drying (w_2). The experiment was conducted in a split-plot format. The process variables are the whole plot factors of the experiment, whereas the mixture components are the sub-plot factors. The data are shown in Table 1.

Table 1. Data for the Vinyl thickness for three-component mixture experiment with two process variable [20].

Obs.	Block	w_1	w_2	s_1	s_2	s_3	y
1	1	−1	−1	1	0	0	10
2	1	−1	−1	0	1	0	4
3	1	−1	−1	0	0	1	3
4	1	−1	−1	0	0.5	0.5	9
5	2	1	−1	0	0	1	7
6	2	1	−1	1	0	0	10
7	2	1	−1	0	1	0	7
8	2	1	−1	0.5	0	0.5	12
9	3	−1	1	0.5	0	0.5	9
10	3	−1	1	0	1	0	5
11	3	−1	1	0.5	0.5	0	8
12	3	−1	1	0	0	1	2
13	4	1	1	0	0.4	0.6	4
14	4	1	1	0.5	0	0.5	7
15	4	1	1	0	1	0	5
16	4	1	1	1	0	0	6
17	5	1	−1	0.5	0.5	0	5
18	5	1	−1	1	0	0	12
19	5	1	−1	0	0	1	16
20	5	1	−1	0	0.6	0.4	9
21	6	−1	1	0	0.5	0.5	11
22	6	−1	1	1	0	0	12
23	6	−1	1	0	0	1	2
24	6	−1	1	0	1	0	9
25	7	1	1	0.5	0.5	0	3
26	7	1	1	0	1	0	5
27	7	1	1	0	0	1	9
28	7	1	1	1	0	0	5

A main effects plus two factor interactions model was assumed for the process variable w_1 and w_2. For the mixture components, the quadratic mixture model was used. The main effects of the process variables were crossed with the linear blending terms only, so the model estimated in [18] is given by

$$y_{ij} = \sum_{i=1}^{3} \beta_i s_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \beta_{ij} s_i s_j + \sum_{i=1}^{2} \alpha_i w_i + \alpha w_1 w_2 + \sum_{i=1}^{2} \sum_{j=1}^{2} \delta_{ij} s_i w_j + \gamma_i + \varepsilon_{ij}$$

They computed the variance components $\sigma^2 = \sigma_\varepsilon^2 + \sigma_\gamma^2$ by R. The $\hat{\sigma}_\varepsilon^2 = 6.764$ and the $\hat{\sigma}_\gamma^2 = 0.00001$. The factor effect estimates will be displayed in the next section to compare it with the proposed methods in this paper.

9. Analysis of the Vinyl Thickness Experiment

The real dataset for the vinyl thickness experiment has been used to apply the Bayesian variable selection approach. The model involves 5 main variables (w_1, w_2, s_1, s_2, s_3) and two factor interaction variables ($w_1 s_1, w_1 s_2, s_1 s_2, s_1 s_3, s_2 s_3, w_2 s_1, w_2 s_2, w_1 w_2$). We used the prior distributions above. Following [21], in a Bayesian framework the final model could be the median probability model consisting of those variables whose posterior inclusion probability $p(\nu_j = 1 | y) \geq 0.5$. The posterior probability of parameter $\beta_j, j = 1, 2, \cdots, p$ being active is approximated by

$$\sum_{q=1}^{its} \frac{\nu_j^{(q)}}{its}, \tag{30}$$

where $\nu_j^{(q)}$ is ν_j sampled at iteration $q = 1, \cdots, its$ of the Metropolis-Hastings within Gibbs sampling algorithm.

We summarise the results of applying the Bayesian variable selection for the data from thickness vinyl experiment. The model which will be used is:

$$y_{ij} = \sum_{i=1}^{3} \beta_i s_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \beta_{ij} s_i s_j + \sum_{i=1}^{2} \alpha_i w_i + \alpha w_1 w_2 + \sum_{i=1}^{2} \sum_{j=1}^{2} \delta_{ij} s_i w_j + \gamma_i + \varepsilon_{ij}$$

The estimates of the 13 parameters of the model have been reported using (SSVS) and (SSVS-SPD) for the response y which is displayed in **Table 1** as well as the estimates by the generalised least estimator (GLS) for comparison purpose.

Figure 2 shows a comparison between SSVS and SSVS-SPD with respect to the resulting approximate posterior probability for the thickness vinyl experiment. The parameters $\beta_{w_1}, \beta_{s_1}, \beta_{s_2}, \beta_{s_3}, \beta_{w_1 s_1}, \beta_{w_1 s_2}$ have the highest posterior probability of being active by both SSVS and SSVS-SPD. This indicates that the six associated variables to these terms play a significant role in this experiment. Followed by these terms, we find the parameters $\beta_{w_1 w_2}, \beta_{s_1 s_2}, \beta_{s_1 s_3}, \beta_{s_2 s_3}$ have an approximate posterior probability of about 0.5 and 0.6 by both SSVS and SSVS-SPD. We note that SSVS and SSVS-SPD tend to consider $\beta_{s_1 s_2}, \beta_{s_1 s_3}, \beta_{s_2 s_3}$ to be significant at an approximate posterior probability of 0.5 and 0.6 while they are not significant by the GLS method. All methods consider $\beta_{w_2}, \beta_{w_2 s_1}, \beta_{w_2 s_2}$ to be non significant with low approximate posterior probability in this experiment.

The Bayesian analysis for the real data of thickness vinyl experiment yielded 10 significant variables. In contrast, the p value for the GLS estimates in Table 2 shows there are 7 significant variables as it excludes the variables associated to the coefficients $\beta_{s_1s_2}, \beta_{s_1s_3}, \beta_{s_2s_3}$. This means that the SSVS and SSVS-SPD yielded in extra 3 significant variables to the model. Table 2 represents the posterior means of the coefficients and standard deviation by both the SSVS and the SSVS-SPD methods and the GLS estimates with the p values for the thickness vinyl experiment. Table 3 shows the posterior mean of the correlation $\hat{\rho}$ and the posterior mean of the total variance $\hat{\sigma}^2$ for both SSVS and SSVS-SPD methods. Figure 3, shows the ACF plots for Markov Chain using Metropolis Hastings within Gibbs sampling algorithm used by SSVS and SSVS-SPD to sample the correlation ρ and the variance σ^2.

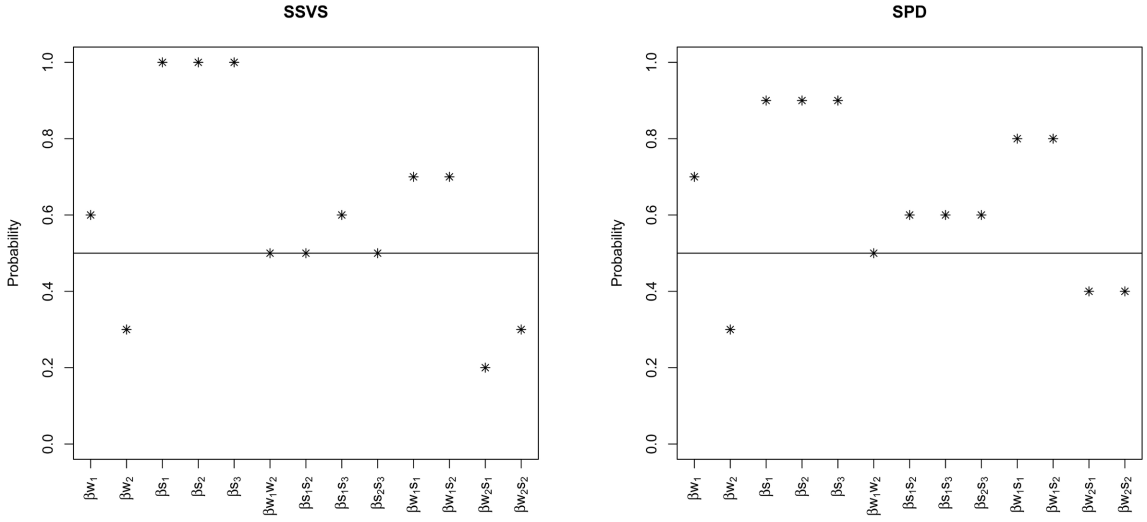

Figure 2. Approximate posterior probability for the thickness vinyl experiment.

Table 2. The estimated coefficients and standard deviations (in parenthesis) for the thickness vinyl experiment by SSVS and SSVS-SPD. The first row is the estimates coefficeients and p-values (in parenthesis) for the GLS.

Method	w_1	w_2	s_1	s_2	s_3	w_1w_2	s_1s_2	s_1s_3	s_2s_3	w_1s_1	w_1s_2	w_2s_1	w_2s_2
GLS	2.7048	−1.4579	9.6836	5.7338	6.0306	−1.2478	−9.2955	5.0613	10.2519	−4.2558	−3.3837	0.5313	1.8096
	(0.006)	(0.1416)	**(0)**	**(0)**	**(0)**	**(0.0189)**	(0.1742)	(0.4597)	(0.0987)	**(0.0038)**	**(0.0208)**	(0.7141)	(0.2275)
SSVS	1.1231	−0.4047	8.7741	5.6278	6.2141	−0.5013	−3.8195	2.9947	3.5948	−2.1752	−1.6525	−0.0646	0.4176
	(1.5031)	(0.8618)	(1.4849)	(1.4178)	(1.3677)	(0.8598)	(5.5208)	(5.0878)	(5.0977)	(2.0511)	(1.8558)	(0.6786)	(0.9723)
SSVS-SPD	1.1101	−0.3762	7.1253	4.1025	4.7905	−0.5875	−3.5483	2.8006	3.2632	−1.9048	−1.4335	−0.0466	0.7020
	(0.9997)	(0.9801)	(2.7811)	(2.5124)	(2.4627)	(1.0781)	(5.5493)	(5.5481)	(5.5821)	(1.9083)	(1.6913)	(1.0416)	(1.2863)

Table 3. Posterior means of the $\hat{\sigma}^2$ and the $\hat{\rho}$ by the SSVS and SSVS-SPD for the thickness vinyl experiment.

Method	$\hat{\sigma}^2$	$\hat{\rho}$
SSVS	11.7622	0.3624
SSVS-SPD	12.3379	0.4906

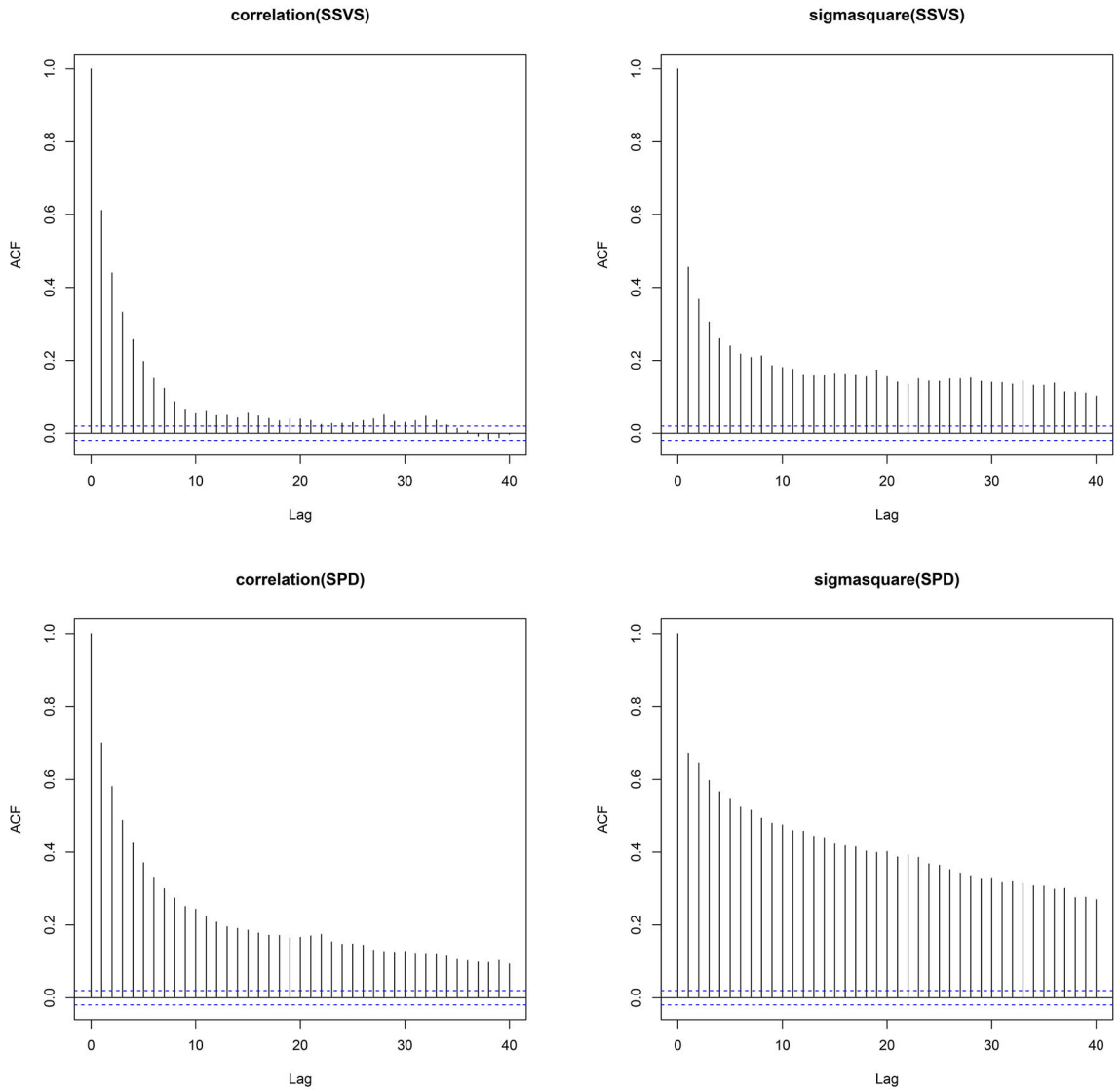

Figure 3. ACF plot for the Markov chain formed by sampling the total variance σ^2 and the correlation ρ by SSVS and SSVS-SPD for the thickness vinyl experiment.

10. Simulation Study Using the Design of the Vinyl Thickness Experiment

We performed a simulation study by generating 1000 datasets, where each dataset would ran for 10000 iterations using MCMC. The SSVS and the SSVS-SPD would be applied at two levels of $\eta = 1$ and $\eta = 10$. We assume that $\sigma_\varepsilon^2 + \sigma_\varepsilon^2 = 10$. Thus, the true value of the total variance $\sigma^2 = 10$. Also, the true value of ρ is 0.5. We will calculate Type I and II error rates using the indicator vector ν and the approximate posterior probability in 30. If the true variable is active but the algorithm yielded a corresponding approximation posterior probability of less than 0.5, this variable would then have Type II error rate. Also, if the true variable is non-active but the algorithm yielded a corresponding approximation posterior probability larger than or equal to 0.5, this variable would

then have Type I error rate. We also will calculate the precision of the point estimates by SSVS and SSVS-SPD by counting the median relative model error (MRME) for the estimates of the SSVS and SSVS-SPD. We focus on the properties of the estimated models by investigating the following properties:

1) consistency in variable selection (frequency in selecting the active/non-active variable), and

2) prediction performance.

For point 1, at 5% significant level, we report Type I error rate (an effect that is truly not significant but the corresponding procedure estimate indicates that it is significant). We also report Type II error rate (an effect that is truly present but the corresponding procedure estimate indicates that it is not significant).

For point 2, following [22] and [23], prediction accuracy is measured by computing the mean-squared error for each penalised estimate $\hat{\boldsymbol{\beta}}_{\lambda}$ as,

$$\mathrm{ME}\left(\hat{\boldsymbol{\beta}}_{\lambda}\right)=\left(\mathbf{X}\hat{\boldsymbol{\beta}}_{\lambda}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{X}\hat{\boldsymbol{\beta}}_{\lambda}-\mathbf{X}\boldsymbol{\beta}\right).$$

The relative model error (**RME**) is the ratio of the model error of the penalised estimates to the model error for the GLS estimates of the fixed effects,

$$\mathbf{RME}=\frac{\mathbf{ME}\left(\hat{\boldsymbol{\beta}}_{\lambda}\right)}{\mathbf{ME}\left(\hat{\boldsymbol{\beta}}_{GLS}\right)},$$

where $\hat{\boldsymbol{\beta}}_{GLS}$ in Equation (13) is the generalised least squares estimator of $\boldsymbol{\beta}$. The median of the relative model error (**MRME**) over 1000 simulated data sets were reported. **MRME** values greater than one indicate that the methods estimates perform worse than the GLS estimates, values near to one indicate that the methods estimates performs in a similar way to the GLS estimates, values less than one indicate that the methods estimates performs better than the GLS estimates.

We perform a simulation study to examine the performance of the SSVS and SSVS-SPD methods. Using the design of the thickness vinyl we generate the response given the true model as:

$$\mathbb{E}\left(\mathbf{Y}\right)=4w_1-3s_1+s_3+4w_1w_2+2s_1s_2-s_1s_3+2w_1s_1+3w_2s_1 \tag{31}$$

Type I and II error rates are displayed in **Table 4** and **Table 5**, for two setting of $\eta=1$ and $\eta=10$. Also, **Table 6** represents the estimated Posterior means of $\hat{\sigma}^2$ and $\hat{\rho}$ by the SSVS and SSVS-SPD from the simulation by using the design of the thickness vinyl experiment. **Figure 4** shows the MRME values at $\eta=1$ and $\eta=10$. We notes that the SSVS at $\eta=1$ have MRME greater than one which indicates that the estimates by the SSVS is worse than the GLS estimates. While at $\eta=10$, the SSVS perform better than the GLS estimates. The SSVS-SPD have similar performance with the GLS estimates at both level of η.

Type II error rates at both level of η generally are low indicating that the active values are easy to detect by both SSVS and SSVS-SPD. Furthermore, Type II error rates by the SSVS-SPD are lower than Type II error rates by the SSVS.

With regard to Type I error rate, for both level of η, the SSVS-SPD have lower Type I error rates than the SSVS. Detecting active variables by both SSVS and SSVS-SPD is better than detecting the non active variables. In summary, we found that the SSVS-SPD detects the active and non-active effect factors with a lower error rate than the SSVS. The analysis of the Median Relative Model Error (MRME) for both SSVS and SSVS-SPD using the thickness vinyl design provides similar behaviour to the GLS except at $\eta = 10$ for the SSVS as detecting active effect factors was harder than the SSVS-SPD.

Table 4. Type I error rate for the simulation by using the design of thickness vinyl experiment.

True non active variable	Method	w_2	s_2	s_2s_3	w_1s_2	w_2s_2
$\eta = 1$	SSVS	0.061	0.054	0.078	0.051	0.042
	SSVS-SPD	0.058	0.037	0.052	0.046	0.040
$\eta = 10$	SSVS	0.045	0.069	0.055	0.077	0.082
	SSVS-SPD	0.050	0.030	0.056	0.069	0.058

Table 5. Type II error rate for the simulation by using the design of thickness vinyl experiment.

True active variable	Method	w_1	s_1	s_3	w_1w_2	s_1s_2	s_1s_3	w_1s_1	w_2s_1
		4	−3	1	4	2	−1	2	3
$\eta = 1$	SSVS	0.004	0.005	0.002	0.005	0.007	0.005	0.004	0.007
	SSVS-SPD	0.003	0.005	0	0.003	0.004	0.003	0.008	0.019
$\eta = 10$	SSVS	0.002	0	0	0.032	0.005	0	0.014	0.005
	SSVS-SPD	0	0	0	0.017	0	0	0.009	0.010

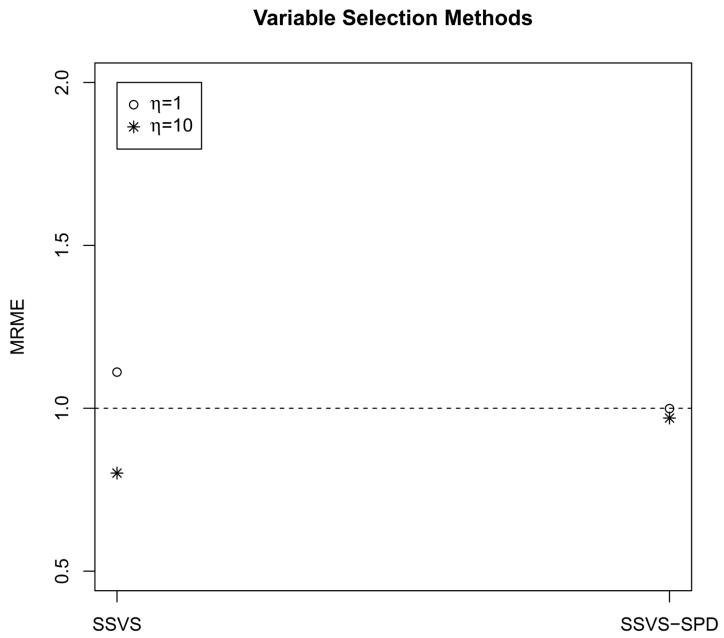

Figure 4. Median relative model error (MRME) for the simulation by using the design of thickness vinyl experiment.

Table 6. Posterior means of the $\hat{\sigma}^2$ and the $\hat{\rho}$ by the SSVS and SSVS-SPD for the simulation by using the design of thickness vinyl experiment.

η	Method	$\hat{\sigma}^2$	$\hat{\rho}$
1	SSVS	10.9	0.56
	SSVS-SPD	11.2	0.50
10	SSVS	9.00	0.62
	SSVS-SPD	9.32	0.68

11. Conclusion

This paper provided an analysis of data from split-plot mixture process experiments using a motivating example from the industrial environment. Specifically, we recommend the use of the SSVS-SPD method for Bayesian variable selection. In our results, we observed that the SSVS-SPD can identify the active variables (linear and two-factors interaction), much better than the SSVS and the traditional used GLS method. However, as expected this comes with the expense of slightly higher Type I error rates. We also observed a better prediction performance for the models chosen by the SSVS-SPD compared to the models chosen by the SSVS method.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Aljeddani, S. (2022) Bayesian Variable Selection for Mixture Process Variable Design Experiment. https://doi.org/10.21203/rs.3.rs-1906056/v1

[2] Tang, X., Xu, X., Ghosh, M. and Ghosh, P. (2016) Bayesian Variable Selection and Estimation Based on Global-Local Shrinkage Priors. *Sankhya A*, arXiv: 1605.07981.

[3] Mitchell, T.J. and Beauchamp, J.J. (1988) Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association*, **83**, 1023-1032. https://doi.org/10.1080/01621459.1988.10478694

[4] George, E.I. and McCulloch, R.E. (1993) Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, **88**, 881-889. https://doi.org/10.1080/01621459.1993.10476353

[5] Geweke. J. (1996) Variable Selection and Model Comparison in Regression. *Bayesian Statistics 5*, Valencia, 5-9 June 1994.

[6] Tan, M.H.Y. and Wu, C.F.J. (2013) A Bayesian Approach for Model Selection in Fractionated Split Plot Experiments with Applications in Robust Parameter Design. *Technometrics*, **55**, 359-372. https://doi.org/10.1080/00401706.2013.778790

[7] George, E.I. and McCulloch, R.E. (1997) Approaches for Bayesian Variable Selection. *Statistica Sinica*, **7**, 339-373.

[8] Aljeddani, S. (2022) Variable Selection in Randomized Block Design Experiment. *Journal of Computational Mathematics*, **12**, 216-231. https://doi.org/10.4236/ajcm.2022.122013

[9] Gilmour, S.G. and Goos, P. (2009) Analysis of Data from Non-Orthogonal Multi-stratum Designs in Industrial Experiments. *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, **58**, 467-484. https://doi.org/10.1111/j.1467-9876.2009.00662.x

[10] Jones, B. and Nachtsheim, C.J. (2009) Split-Plot Designs: What, Why, and How. *Journal of Quality Technology*, **41**, 340-361. https://doi.org/10.1080/00224065.2009.11917790

[11] Aljeddani, S. (2019) Statistical Analysis of Data from Experiments Subject to Restricted Randomisation. University of Southampton, Southampton.

[12] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2014) Bayesian Data Analysis. Vol. 2, CRC Press, Boca Raton. https://doi.org/10.1201/b16018

[13] O'Hagan, A. and Forster, J. (2004) Kendall's Advanced Theory of Statistics: Bayesian Statistics. Vol. 2B, Arnold, London.

[14] Gilks, W.R., Richardson, S. and Spiegelhalter, D. (1995) Markov Chain Monte Carlo in Practice. CRC Press, Boca Raton. https://doi.org/10.1201/b14835

[15] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**, 1087-1092. https://doi.org/10.1063/1.1699114

[16] Hastings, W.K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**, 97-109. https://doi.org/10.1093/biomet/57.1.97

[17] Gelman, A., Roberts, G.O., Gilks, W.R., *et al*. (1996) Efficient Metropolis Jumping Rules. *Bayesian Statistics*, **5**, 599-607.

[18] Cornell, J.A., Kowalski, S.M. and Vining, G.G. (2002) Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables. *Technometrics*, **44**, 72-79. https://doi.org/10.1198/004017002753398344

[19] Cornell, J.A. (2011) Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data. John Wiley & Sons, Hoboken.

[20] Goos, P. and Donev, A.N. (2007) Tailor-Made Split-Plot Designs for Mixture and Process Variables. *Journal of Quality Technology*, **39**, 326-339. https://doi.org/10.1080/00224065.2007.11917699

[21] Barbieri, M.M., Berger, J.O. *et al*. (2004) Optimal Predictive Model Selection. *The Annals of Statistics*, **32**, 870-897. https://doi.org/10.1214/009053604000000238

[22] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[23] Ibrahim, J.G., Zhu, H., Garcia, R.I. and Guo, R. (2011) Fixed and Random Effects Selection in Mixed Effects Models. *Biometrics*, **67**, 495-503. https://doi.org/10.1111/j.1541-0420.2010.01463.x