Comparison of Some Spike-and-Slab Priors for Bayesian Variable Selection in Multiple Linear Regression

G.M. Oyeyemi[1*]; Y.A. Olanrewaju[1]; and R.O. Kolawole[2]

[1]Department of Statistics, University of Ilorin, Ilorin, Nigeria.
[2]Department of Building Technology, The Federal Polytechnic Offa, Offa, Nigeria.

E-mail: gmoyeyemi@gmail.com*

ABSTRACT

Variable selection has been a very essential challenge in building a multiple regression model. Exclusion of influential covariates or including covariate with zero effect will no doubt affect the estimation precision and as well the predictive accuracy of the model. “Spike-and-Slab prior” is an increasingly popular variable selection approach used in the Bayesian framework, which aids the variable selection and the estimation of regression parameters.

In this research, the performances of the MCMC implementation for some versions of spike and slab priors for variable selection in normal linear regression models were investigated with regards to posterior inclusion probability for the simulated data under different setting (independent and correlated covariates, difference variance scales and varying sample sizes). Evidence from the simulation study revealed that the selected priors have similar performance under the independent setup and correlated setup, but the standard errors of coefficient estimates are higher for correlated covariates compare to independent covariates. The mean estimates of the coefficients get closer to the true coefficient values as the sample size increases under all different priors considered, and also the posterior inclusion probability depends on the size of variance of the slab component.

(Keywords: covariates, posterior distribution, precision, regression coefficients, spike-and-slab priors)

INTRODUCTION

Regression model is among the most popular statistical methods employ to investigate the impact of independent variable on the response variable. In the normal linear regression model, the mean of the dependent variable is assumed to be a linear function of the influential covariates. In the availability of potentially large set of observed independent variables, the inclusion of non-influential variables in the model or exclusion of important variables from the model may affect the predictive accuracy of the model. Therefore, methods of variable selection are indisputably required to correctly distinguish between zero effects and non-zero effects variables.

In the Bayesian setup, variable selection is accomplished by assigning prior distributions to the coefficients of the independent variables. Although the choice of this prior is assumed to be specified by the nature of data to be analyzed, but a particular method called ‘Spike-and-Slab’ priors have gained a widespread attention. These priors are mixture of two distributions, one with mass centered at zero (Spike) and the other with mass spread over a wide range of non-zero values (Slab).

This type of priors was proposed by Mitchell and Beauchamp (1988) in order to facilitate variable selection. Spike-and-Slab priors are actually put forward for the purpose of variable selection. The probability that a covariate is included in the final model is known as the posterior inclusion probability. It can simply be estimated by the mean value of the total number of posterior samples. Variable selection then depends on the posterior distribution of the indicator variable which is estimated by the empirical frequency of 1 and 0, respectively. The higher the posterior mean of the indicator variable, the higher is evidence that the coefficient of a certain covariate is different from zero and thus have a significant impact on the dependent variable.

In this paper, the performance of some selected different version of spike and slab priors on variable selection was investigated. Priors where the spike is a discrete point mass at zero and the slab has a normal distribution (Independent-priors and Zellner's g-prior) are considered. A second type of priors considered are Stochastic Search Variable Selection (SSVS)-prior and Normal Mixtures of Inverse Gamma (NMIG)-prior, where both the spike and the slab have continuous distributions.

Variable selection actually helps to distinguish between influential and non-influential variables. In a real data set, it is very rare, that the true regression coefficients are either zero or large; the sizes are more likely to be tapered towards zero. (O'Hara and Sallanpaa, 2009).

The mixture of priors for Bayesian variable selection in the linear regression models was originally proposed by Mitchell and Beauchamp (1988) and was remarkably made popular by George and McCulloch (1993, 1997) and Smith and Kohn (1996) among others. The method was extended to multivariate linear regression model by Brown et al. (1998, 2002). The appropriate reviews of some features of the selection priors can be found in Chipman et al. (2001) and Clyde and George (2004).

A prior of the type $\beta\gamma|\sigma 2 \sim N(0, c(X' \gamma X\gamma)^{-1}\sigma^2)$ known as Zellner's g-prior, which uses the design matrix of the current sample was proposed by Smith and Kohn (1996). Liang et al. (2008) and Cui and George (2008) have investigated formulations that use a complete Bayesian approach by imposing mixtures of g-priors on *c*. They also proposed hyper-g priors for *c* which leads to closed form marginal likelihoods and nonlinear shrinkage through Empirical Bayes procedures.

George and McCulloch (1993) proposed and develop a procedure that uses probability considerations for selecting promising subsets, the procedure consists embedding the regression setup in a hierarchical normal mixture model where latent variables are used to identified subset choices, and the promising subsets of predictors is identified as those with higher posterior probability.

Walli and Wagner (2011) compare the MCMC implementations for several spike and slab priors with regard to posterior inclusion probabilities and their sampling efficiency for simulated data. They further investigate posterior inclusion probabilities analytically for different slabs in two simple settings. And the application was illustrated on a data set of psychiatric patients where the goal was to identify covariates affecting metabolism.

In a situation where it is not feasible to makes the full exploration of the model space due to the large number of covariates, the Monte Carlo Markov Chain methods can be used as random search to quickly and efficiently explore the models with high posterior probability (George and McCulloch,1997).

Variable selection can be efficiently achieved by setting a threshold to the marginal posterior probabilities of inclusion. Barbieri and Berger (2002) put forward that the median probability model is the model that includes those covariates having posterior inclusion probability of at least "0.5".

MATERIALS AND METHODS

Model Specification

Consider the following linear model;

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ . \ . \ . + \beta_k x_k \ + \varepsilon \quad \textbf{(1)}$$

Where y is the dependent variable and $x_1, \ldots, x_k$are the independent variables. The $\boldsymbol{\varepsilon}$ is error term, and it assume to be independent and identically distributed normal random variables with variance $\sigma^2 = h^{-1}$.

This yields the distribution of the regression model:

$$\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{X\beta}, \boldsymbol{h}^{-1}.\boldsymbol{I}) \qquad \textbf{(2)}$$

Where ***I*** denotes the unit matrix of dimension n.

For Bayesian regression model with large k covariates, it is quite a computational challenge to consider all models with possible subsets of covariates. Therefore, this can be addressed by indicator random variable approach with high posterior probability using MCMC methods.

A new indicator variable δ_j for each covariates coefficient β_j is defined, where δj= 0 represents exclusion and δj= 1 represents inclusion of the regressor in the model.

$$\delta_j = \begin{cases} 0, if\ \beta_j = 0 \\ 1, otherwise \end{cases}$$

The linear regression model above can be express in matrix form as stated below:

$$y = X\beta + \varepsilon \qquad \textbf{(3)}$$

where **y** is the N×1 vector of the response variable, **X** is the N×k design matrix, ***β*** are vector of regression coefficients including the intercept, and ε is the N×1 error vector.

Hence, by applying the indicator variable to each covariate coefficient, the regression model becomes:

$$y_i = \beta_0 + \delta_1\beta_1 x_{i1} + \delta_2\beta_2 x_{i2} + \ldots + \delta_k\beta_k x_{ik} + \varepsilon_i \qquad \textbf{(4)}$$

The $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_k)$ is the vector that determines which elements of $\boldsymbol{\beta}_l$ are to be restricted to zero or included in the model. Once the value $\boldsymbol{\delta}$ is chosen, the following reduced regression model in matrix form is obtained:

$$y = X^{\delta}\beta^{\delta} + \varepsilon \qquad \textbf{(5)}$$

where β^{δ} contains only the nonzero elements of $\boldsymbol{\beta}$ and the design matrix X^{δ} contains only the columns of **X** corresponding to nonzero effects. The intercept does not have an indicator variable as β_0 is by all means included in the model.

Prior Specification

From (5) above, the model parameters are; the model indicator $\boldsymbol{\delta}$, the regressor effects $\boldsymbol{\beta}$, and the error variance $\boldsymbol{\sigma^2}$. In the Bayesian framework, the goal is to derive the posterior density of all the parameters, i.e.,

$$p(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\sigma^2} | \mathbf{y}) \ \alpha p(\mathbf{y} | \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\sigma^2})\ p(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\sigma^2})$$

To specify a prior distribution, we assume that it has the structure:

$$p(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\sigma^2}) = p(\boldsymbol{\sigma^2}, \boldsymbol{\beta} | \boldsymbol{\delta}) \prod p(\boldsymbol{\delta}_j)$$

As δj is a binary variable, a straightforward choice of a prior for δj is:

$$p(\delta_j = 1) = \pi, \qquad j = 1, \ldots, k..$$

where π is a fixed inclusion probability between 0 and 1. For σ^2 and those effects of β which are not restricted to be zero, β^{δ}, conjugate priors are used, i.e.,

$$\sigma^2 \quad \sim \quad G^{-1}(s_0, S_0)$$

$$\boldsymbol{\beta}^{\boldsymbol{\delta}} | \sigma^2, \boldsymbol{\delta} \quad \sim \quad N(b_0^{\delta}, B_0^{\delta}\sigma^2)$$

For the prior inclusion probability $p(\delta j = 1) = \pi$, we use a hierarchical prior: $\pi \sim \beta(c_0, d_0)$

Spike-and-Slab Priors

Priors where the spike is a discrete point mass at zero and the slab has a normal distribution such as independent-priors and Zellners' g-prior are considered. A second type of priors considered are SSVS-prior and NMIG-prior where both the spike and the slab have continuous distributions. introducing the indicator variable δ_j in the linear regression model, the resulting prior for the regression coefficients is an example of a spike and slab prior. More formally the prior can be written as:

$$p(\beta_j | \delta_j) = \delta_j p_{slab}(\beta_j) + (1 - \delta_j)\ p_{spike}(\beta_j)$$

The coefficient is assumed to belong to either the spike distribution or the slab distribution, this depends on the value of δj.

Independent Prior

In the simplest case of a non-informative prior for the coefficients each regressor effect is assumed to follow the same distribution and to be independent of the other regressors:

$$p(\beta^{\delta} | \sigma^2) = N(\mathbf{a}_0^{\delta}, \mathbf{A}_0^{\delta}\sigma^2) = N(\mathbf{0}, c\sigma^2\mathbf{I})$$

where c is an appropriate chosen constant.

Zellner's g-Prior

Like the independence prior the g-prior assumes that the effects are a priori centered at zero, but the covariance matrix $\mathbf{A}_0$ is a scalar multiple of the Fisher information matrix, thus taking the dependence structure of the covariates into account:

$p(\beta^{\delta}|\sigma^2) = N(\mathbf{a}_0^{\delta}, \mathbf{A}_0^{\delta}\sigma^2) = N(\mathbf{0}, g((\mathbf{X}^{\delta})'(\mathbf{X}^{\delta}))^{-1}\sigma^2)$

Stochastic Search Variable Selection (SSVS) Prior

Formally, the prior construction is given as follows:

$\beta_j|v_j = v_j N(0, \varphi^2) + (1-v_j) N(0, c\varphi^2)$

$$v_j = \begin{cases} c, & p(c) = 1 - w \\ 1, & p(c) = w \end{cases}$$

$w \sim B(c_0, d_0)$

$p_{spike}(\beta_j) = N(0, c\varphi^2)$

$p_{slab}(\beta_j) = N(0, \varphi^2)$

where c is a very small constant and ψ^2 is a fixed value chosen large enough to cover all reasonable values (Liang et al., 2008).

Normal Mixture of Inverse Gamma (NMIG) - Prior

$\beta_j|v_j = v_j N(0, \varphi_j^2) + (1-v_j) N(0, c\,\varphi_j^2)$

$\varphi_j^2 \sim \quad G^{-1}(a\varphi_0, b\varphi_0)$

$$v_j = \begin{cases} c, & p(c) = 1 - w \\ 1, & p(c) = w \end{cases}$$

$w \sim B(c_0, d_0)$

The resulting prior for the variance parameter is a mixture of scaled inverse Gamma distribution:

$p(\varphi_j^2) = (1 - \omega)G^{-1}(\varphi_j^2|a\psi 0, s0b\psi 0) + \omega G^{-1}(\varphi_j^2|a\psi 0, s1b\psi 0)$

MCMC Scheme for Both Independent Prior and Zelner's g-Prior

(1) sample each element δ_j of $\boldsymbol{\delta}$ separately from $p(\delta_j \mid \delta\backslash_j, \mathbf{y}) \propto p(\mathbf{y}| \delta_j, \delta\backslash_j)p(\delta_j, \boldsymbol{\delta}\backslash_j)$, where, $\delta\backslash_j$ denotes the vector $\boldsymbol{\delta}$ without element δ_j.

(2) sample $\sigma 2|\boldsymbol{\delta}$ from $G^{-1}(s_N^{\delta}, S_N^{\delta})$

(3) sample the intercept μ from $N(\bar{y}, \sigma^2/N)$

(4) sample the nonzero elements $\beta_j|\sigma 2$ in one block from $N(\mathbf{a}_N^{\delta}, \mathbf{A}_N^{\delta}\, \sigma^2)$

The marginal likelihood of the data conditioning only on the indicator variables is given as:

$$p(\mathbf{y}|\boldsymbol{\delta}) = \frac{1}{(2\pi)^{N/2}} \frac{|A_N^{\delta}|^{1/2}}{|A_0^{\delta}|^{1/2}} \frac{\Gamma(S_N) S_0^{s0}}{\Gamma(s_0)(S_N^{\delta})^{s_N}}$$

with the posterior moment:

$\mathbf{A}_N^{\delta} = ((\mathbf{X}^{\delta})'\mathbf{X}^{\delta} + (\mathbf{A}_0^{\delta})^{-1})^{-1}$

$\mathbf{a}_N^{\delta} = \mathbf{A}_N^{\delta}\, ((\mathbf{X}^{\delta})'\mathbf{y} + (\mathbf{A}_0^{\delta})^{-1}\mathbf{a}_0^{\delta})$

$s_N = s0 + (N - 1)/2$

$S_N = S0 + (\mathbf{y} - \mathbf{X}\mathbf{A}_0^{\delta})'(\mathbf{y} - \mathbf{X}\mathbf{A}_0^{\delta})/2 + (\boldsymbol{\beta} - \mathbf{b}0)'(\mathbf{A}_0^{\delta})^{-1}(\mathbf{A}_0^{\delta} - \mathbf{a}_0^{\delta})/2$

Under the Zelner's g-prior, the marginal likelihood is given as; $S(X^{\delta})$

$$p(\mathbf{y}|\boldsymbol{\delta}) = \frac{(1+g)^{-q\delta/2}}{\sqrt{N}(\pi)^{(N-1)/2}} \frac{\Gamma(\frac{N-1}{2})}{S(X\delta)^\wedge(N-1)/2}$$

where,

$$S(X\delta) = \frac{\|y\|^2}{1+g}\left(1 + g(1 - R(X\delta)^2)\right)$$

δ is the number of nonzero elements in $\boldsymbol{\delta}$ and $R(\mathbf{X}\delta)$ is the coefficient of determination.

The posterior moments are given as:

$$\mathbf{A}_N^{\delta} = \frac{g}{1+g} (\mathbf{X}^{\delta\prime}\mathbf{X}^{\delta})^{-1}$$

$$a_N^{\delta} = = \frac{g}{1+g} (\mathbf{X}^{\delta\prime}\mathbf{X}^{\delta})^{-1}\, \mathbf{X}^{\delta\prime} y$$

$$s_N = s0 + (N - 1)/2$$

$$S_N = S0 + S(X\delta)/2$$

MCMC Scheme for SSVS Prior

(1) sample the intercept μ from $N(\bar{y}, \sigma^2/N)$

(2) sample each Vj, j=1,…,k. from $p(Vj| \beta_j, w)$

(3) sample w from $B(c_0 + n_1, d_0 + k - n_1)$

(4) sample β_j from $N(\mathbf{a_N}, \mathbf{A}_N)$, where $(\mathbf{A}_N)^{-1} = X'X/\sigma^2 + D^{-1}$, $\mathbf{a_N} = \mathbf{A}_N X'y/\sigma^2$ and $D = \text{diag}(\varphi^2 Vj)$

(5) sample $\sigma 2|\boldsymbol{\delta}$ from $G^{-1}(s_N, S_N)$

MCMC Scheme for NMIG-Prior

(1) sample the common μ from $N(\bar{y}, \sigma^2/N)$

(2) for each j=1,…,k, sample Vj from $p(Vj| \beta_j, \varphi^2_j, w, y)$

(3) for each j=1,…,k, sample φ^2_j from $p(\varphi^2_j \,|\, \beta_j, Vj$

(4) sample w from $B(c_0 + n_1, d_0 + k - n_1)$, where $n1 = \sum_j I_{(vj=1)}$

(5) sample the error variance σ^2 from $G^{-1}(s_N, S_N)$

ANALYSIS AND RESULTS

The data employed for this study were simulated from the environment of R statistical package. The study employed MCMC (Markov Chain Monte Carlo) experiment to empirically investigate the performances of the selected four Spike-and-Slab priors for Bayesian variable selection with regard to posterior inclusion probabilities.

Two setups were considered; the independent and also correlated covariates. Data sets with nine predictor variables each with 50 observations and response variable were generated with the following setting: The nine covariates are drawn from multivariate normal distribution with mean vector equal to Zero and covariance matrix equal to the identity matrix. And the response variable Y is simulated in accordance with the model;

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_9 x_{i9} + \varepsilon_i$$

Where the intercept is set to 1 and the error term is generated from the $N(0,1)$ distribution for all data sets. Three different sample sizes n = (50, 500 and 1000) relating to small, medium, and large sample sizes were adopted to examine the effect of sample size on the performance of the different spike and slab priors under study.

For the correlated covariates, the setup configuration of Tibshirani (1996) is employed where the nine covariates are drawn from multivariate normal distribution with the mean vector equal to Zero and variance covariance matrix is set to be:

$\sum_{ij} = cor(x_i, x_j) = \rho^{|i-j|}$. Where ρ=0.8.

In order to investigate the correct selection of the influential covariates, following the approach of Walli and Wagner (2011), three regression effects were set for each of the value of the coefficients for both the independent and the correlated covariates, where "2" is assigned as "Strong effects", "0.2" as "Weak effects", and "0" as "Zero effects". This is done such that the coefficient vector for independent covariates is set to be:

β'=(2, 2, 2, 0.2, 0.2, 0.2, 0, 0, 0)

And for the simulation with correlated covariates, the coefficient vector is set to be:

β'=(2, 2, 0, 2, 0.2, 0, 0, 0.2, 0.2)

For each data set, MCMC is run for M=1000 iteration without burn in, and the coefficient estimation and variable selection is performed jointly.

Sample Size Effect

The estimated coefficients and their respective standard deviations for each of the different priors under study are displayed in Table1. The results show that at n=50, the mean of the estimate is close to the true coefficient values for both strong and zero effects but smaller for weak effect, whereas the same result was also recorded at n=500 but with a little improvement to weak effect. Finally, at n=1000, the mean estimates of all the covariates are close to their respective true coefficient values.

Table1: Simulation Results for Different Priors under Different Sample Sizes.

Sample Size		Independent Prior		Zellner's g prior		NMIG prior		SVSS prior	
		Coef.	*Std. dev*	*Coef.*	*Std. dev*	*Coef.*	*Std. dev*	*Coef.*	*Std. dev*
$n = 50$	$\beta_1 = 2$	1.9376	0.2020	1.8787	0.1759	2.1101	0.1523	1.8121	0.2335
	$\beta_2 = 2$	2.2178	0.1932	2.0077	0.1620	1.6778	0.1681	1.8385	0.2342
	$\beta_3 = 2$	1.8596	0.2028	2.0615	0.1875	2.0132	0.1473	2.3316	0.2148
	$\beta_4 = 0.2$	0.0013	0.0222	0.0005	0.0240	0.0190	0.0637	0.1848	0.1657
	$\beta_5 = 0.2$	0.0032	0.0414	0.0839	0.2196	0.0287	0.0783	0.1794	0.1665
	$\beta_6 = 0.2$	0.0043	0.0394	0.0007	0.0201	0.0160	0.0739	0.1752	0.2257
	$\beta_7 = 0$	0.0033	0.0327	0.0005	0.0094	0.0413	0.0757	0.1418	0.2068
	$\beta_8 = 0$	0.0043	0.0383	0.0099	0.0586	0.0060	0.0624	0.1026	0.1836
	$\beta_9 = 0$	0.0008	0.0164	0.0010	0.0169	0.0186	0.0714	0.1146	0.1893
$n = 500$	$\beta_1 = 2$	1.9377	0.0464	1.9743	0.0385	1.9669	0.0453	2.0215	0.0546
	$\beta_2 = 2$	1.9936	0.0429	1.9672	0.0373	1.9915	0.0436	1.9976	0.0542
	$\beta_3 = 2$	2.0078	0.0472	2.0597	0.0400	2.0282	0.0469	1.9686	0.0528
	$\beta_4 = 0.2$	0.2279	0.0448	0.2314	0.0408	0.1520	0.0426	0.2673	0.0500
	$\beta_5 = 0.2$	0.057	0.0725	0.2350	0.0393	0.1251	0.0428	0.1310	0.0523
	$\beta_6 = 0.2$	0.2200	0.0453	0.2322	0.0408	0.1133	0.0425	0.0834	0.0529
	$\beta_7 = 0$	0.0001	0.0038	0.0007	0.0094	0.0195	0.0427	0.0535	0.0519
	$\beta_8 = 0$	0.0000	0.0039	0.0073	0.0243	0.0086	0.0421	0.0619	0.0574
	$\beta_9 = 0$	0.0004	0.0049	0.0020	0.0124	0.0225	0.0411	0.0545	0.0549
$n = 1000$	$\beta_1 = 2$	2.0003	0.0382	1.9953	0.0309	1.9642	0.0379	1.9663	0.0478
	$\beta_2 = 2$	1.9899	0.0391	2.0369	0.0318	2.0240	0.0400	1.9780	0.0473
	$\beta_3 = 2$	1.9912	0.0382	1.9575	0.0314	1.9958	0.0398	1.9172	0.0459
	$\beta_4 = 0.2$	0.2398	0.0391	0.2341	0.0319	0.2420	0.0336	0.1028	0.0466
	$\beta_5 = 0.2$	0.1973	0.0413	0.2343	0.0332	0.1410	0.0363	0.2488	0.0427
	$\beta_6 = 0.2$	0.1562	0.0552	0.1757	0.0337	0.2163	0.0348	0.1766	0.0450
	$\beta_7 = 0$	0.0000	0.0026	0.0033	0.0138	0.0373	0.0368	0.0108	0.0487
	$\beta_8 = 0$	0.0003	0.0057	0.0010	0.0091	0.0101	0.0388	0.0233	0.0459
	$\beta_9 = 0$	0.0001	0.0032	0.0001	0.0064	0.0239	0.0376	0.0100	0.0477

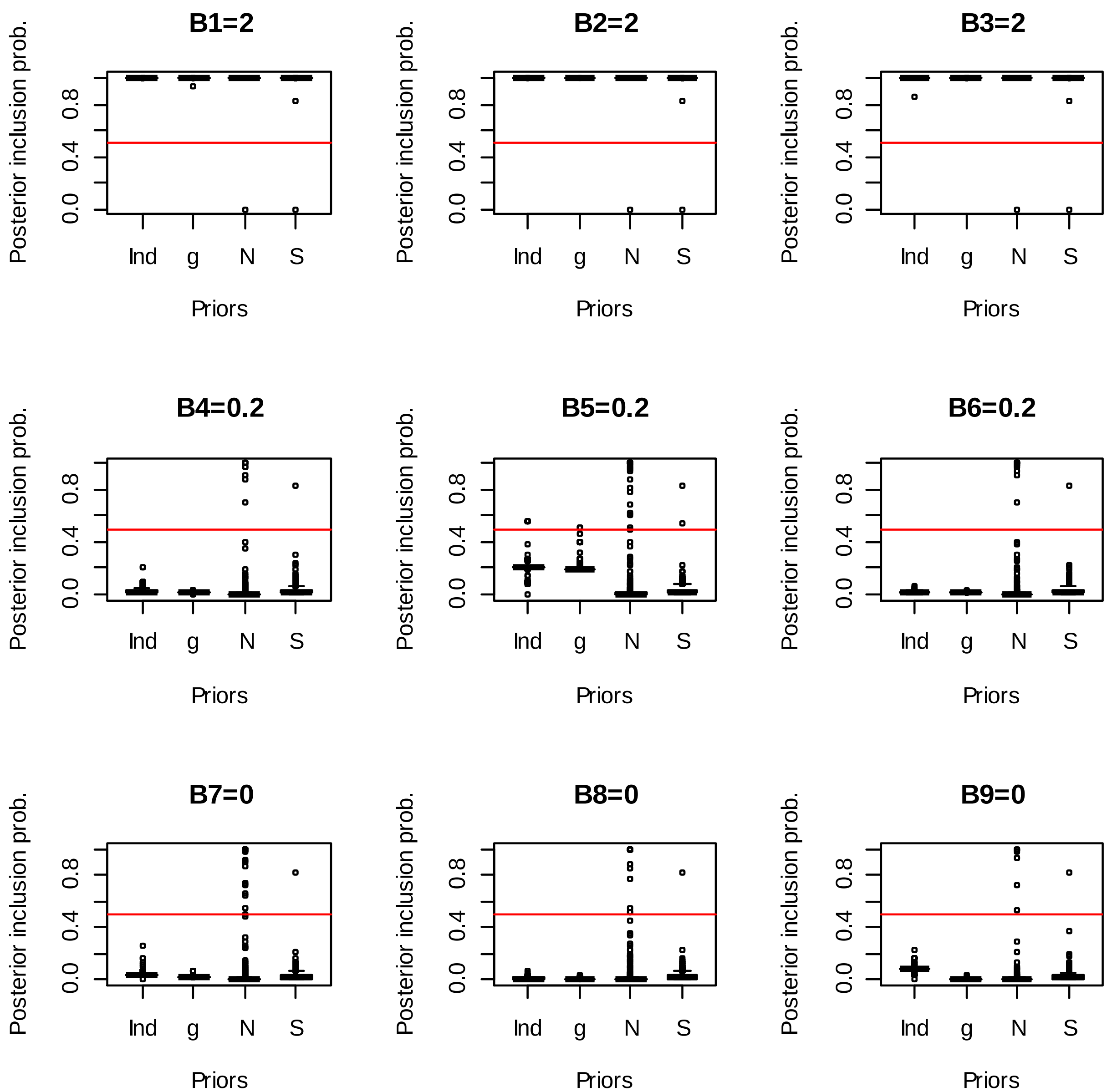

Figure 1: Independent Covariates: Plots of the Posterior Inclusion Probabilities.

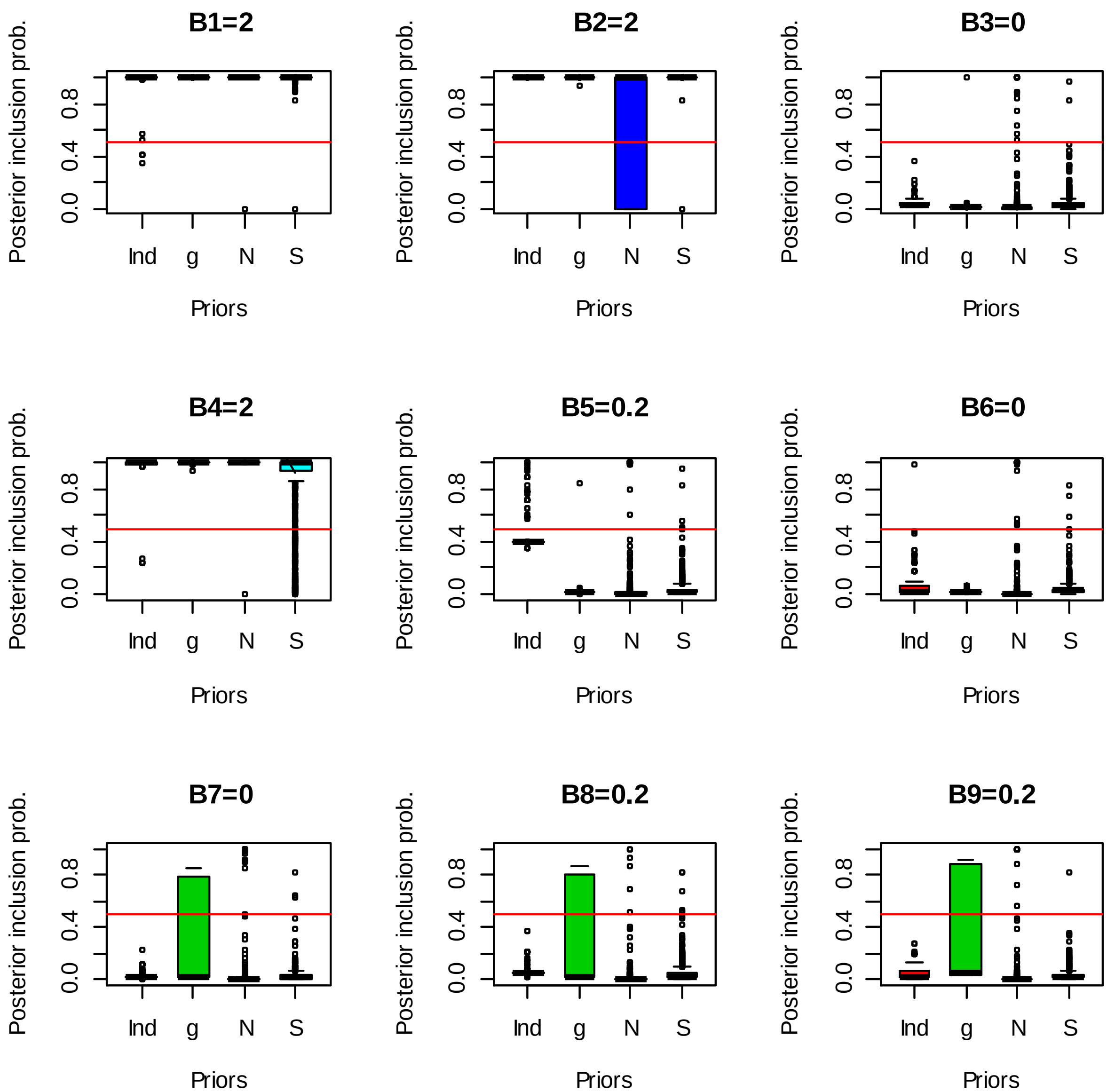

Figure 2: Correlated Covariates: Plots of the Posterior Inclusion Probabilities.

Performances of Variable Selection

Procedures: For the strong covariates, the inclusion probabilities are equal to one under the four selected priors for the two groups of variance components. This implies that the strong coefficients are sampled from only the slab components of the priors, for any chosen variance size either large or small. However, for weak coefficients, the choice of the size of prior variance has impact on the inclusion probabilities. If the size of prior variance is large, the inclusion probabilities are low, and if the size of prior variance is small, the inclusion probabilities are slightly high. This implies that the lower the size of the prior variance the higher the inclusion probabilities. For the zero effect, the similar result is obtained as those of weak coefficient.

As for the correlated covariates, the results obtained are actually very similar to those obtained under independents covariates.

It is observed that errors have increased compare to those of independent covariates. This is as a result of presence of collinearity, yielding large estimation errors.

DISCUSSION OF RESULTS

The performance of the four different selected most popular spike and slab priors in the literature were compared, which are, Independent prior, Zellner's g-prior, SVSS-prior and NMIG-prior. The aim of this study is to show or otherwise, that the four selected priors perform differently due to different location and variance parameters under different scenarios. The results obtained from the simulation study are as follows:

The four selected priors performed similar under the same sample size. At n=50, the mean of the estimate is close to the true coefficient values for both strong and zero effects but smaller for weak effect, whereas the same result was also recorded at n=500 but with a little improvement to weak effect. Finally at n=1000, the mean estimates of all the covariates are close to their respected true coefficient values. For independent covariates, it was revealed that, for strong effect and zero effect, the mean of the estimate is close to the true coefficient, while it is smaller for weak effect. In line with this result, it can be concluded that a large prior variance caused negligible shrinkage.

Also the result of the small variance group is somewhat similar to that of the large variance group but its square error (*SE*) is a little bit higher than that of large variance group most especially for strong effect. It can also be observed that independent prior, g-prior and NMIG prior performed rather similar, with low square error under large component group and higher square error under small variance component group. Whereas SVSS-prior has lower square error under small variance group compare to large variance group. The results also reveal that the strong coefficients are sampled from only the slab components of the priors, for any chosen variance size either large or small. If the size of prior variance is large, the inclusion probabilities are low, and if the size of prior variance is small, the inclusion probabilities are slightly high.

For correlated covariates, the results obtained are actually very similar to those obtained under independents covariates. Under all priors, the mean estimates for strong and zero effect are close to the true values while the estimates of the weak effects are underestimated. It is observed that errors have increased compare to those of independent covariates. This is as a result of presence of collinearity, yielding large estimation errors. The estimation accuracy is not significantly affected by the correlation among the covariates.

CONCLUSIONS

In general, it can be concluded that: all the four priors considered in this study have quite similar results under the same scale setting of parameters of the prior variances. The mean estimates of the coefficients gets closer to the true coefficient values as the sample size increases under all different priors. As the variance's size increases, the posterior inclusion probability for weak and zero effect decreases. Thus, posterior inclusion probability depends on the size of variance of the slab component. The priors are unable to differentiate between weak effect and zero effect. Both effects are either simultaneously included or excluded in the model. The selected priors perform similar under both Independent and Correlated covariates setup. However, the standard errors of coefficient estimates are higher for correlated covariates compare to independent covariates.

REFERENCES

1. Barbieri, M. and O.J. Berger. 2004. "Optimal Predictive Model Selection". *Annals of Statistics*. 32(3): 870 – 897.

2. Brown, P.J., M. Vannucci, and T. Fearn. 2002. "Bayes Model Averaging with Variable Selection of Regression". *Journal of the Royal Statistical Society, Series B*. 64(3): 519 –536.

3. Carlin, B.P. and S. Chib. 1995. "Bayesian Model Choice via Markov Chain Monte Carlo Methods". *Journal of the Royal Statistical Society, Series B*. 57: 473 – 484.

4. Chipman, H., E. George, and R. McCulloch. 2001. "The Practical Implementation of Bayesian Model Selection". *Model Selection, IMS*. 67–116.

5. Clyde, M. and E. George. 2004. "Model Uncertainty". *Statistical Science*. 19(1): 81–94.

6. Cui, W. and E. George. 2008. "Empirical Bayes vs. Fully Bayes Variable Section". *Journal of the Statistical Planning and Inference*. 138: 888 – 900.

7. George, E. and R. McCulloch. 1993. "Variable Selection via Gibbs-Sampling". *Journal of the American Statistical Association*. 85: 398–409.

8. George, E. and R. McCulloch. 1997. "Approaches for Bayesian Variable Selection". *Statistica Sinica*. 7: 339 – 373.

9. Geyer, C.J. 1992. "Practical Markov Chain Monte Carlo". *Statistical Science*. 7(4): 473–511.

10. Greene, W. 2000. *Econometric Analysis, fourth edition*. Prentice-Hall: Upper Saddle River, NJ.

11. Ishwaran, H. and J. Rao. 2003. "Detecting Differentially Expressed Genes in Microarrays using Bayesian Model Selection". *Journal of the American Statistical Association*. 98(1): 438 – 455.

12. Kass, R.E. and A.E. Raftery. 1995. "Bayes Factors". *Journal of the American Statistical Association.* 90(1): 773–795.

13. Koop. G. 2003. *Bayesian Econometrics*. John Wiley & Sons Ltd. New York, NY.

14. Ley, E. and M.F.J. Stee. 2007. "Jointness in Bayesian Variable Selection with Application to Growth Regression". *Journal of Macroeconomics*. 29(1): 476–493.

15. Leamer, E. 1978. "Regression Selection Strategies and Revealed Priors". *Journal of the American Statistical Association*. 73: 580–587.

16. Liang, F., R. Paulo, G. Molina, C.A. Clyde, and J.O. Berger. 2008. "Mixtures of g Priors for Bayesian Variable Selection". *Journal of the American Statistical Association*. 103(1): 410–423.

17. Malsiner-Walli, G. 2010. "Bayesian Variable Selection in Normal Regression Models". Unpublished Master's Thesis, Johannes Kepler Universitat, Institut fur Ange- ¨wandte Statistik: Linz.

18. Mitchell, T.J. and J.J. Beauchamp. 1988. "Bayesian Variable Selection in Linear Regression". *Journal of the American Statistical Association*. 83: 1023–1036.

19. O'Hara, R.B. and M.J. Sillanpaa. 2009. "Review of Bayesian Variable Selection Methods: What, How and Which". *Bayesian Analysis*. 4(1): 85–118.

20. Raftery, A., D. Madigan, and J. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models". *Journal of the American Statistical Association*. 92(437): 179–191.

21. Tibshirani, R. 1996. "Regression Shrinkage and Selection via Lasso". *Journal of the Royal Society B*. 58(1): 267–288.

22. Zellner, A. 1986. "On Assessing Prior Distributions and Bayesian Regression Analysis with G-Prior Distributions". *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*. 90(1): 233–243.

SUGGESTED CITATION

Oyeyemi, G.M. Y.A. Olanrewaju, and R.O. Kolawole. 2019. "Comparison of Some Spike-and-Slab Priors for Bayesian Variable Selection in Multiple Linear Regression". *Pacific Journal of Science and Technology*. 20(1):131-140.

[Pacific Journal of Science and Technology]