



Sensitivity of the Posterior Mean on the Prior Assumptions: An Application of the Ellipsoid Bound Theorem

Olawale B. Akanbi^{1*}, Olusanya E. Olubusoye¹ and Oluwaseun O. Odeyemi¹

¹Department of Statistics, University of Ibadan, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author OBA performed part of the statistical analysis and wrote part of the literature searches. Author OEO designed the study and the protocol. Author OOO wrote the first draft of the manuscript, managed the other part of the analyses of the study and the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2020/v26i730291

Editor(s):

(1) Dr. Patricia J. Y. Wong, Nanyang Technological University, Singapore.

Reviewers:

(1) Meghana Kshirsagar, University of Limerick, Ireland.

(2) Cheruiyot Nickson, Egerton University, Kenya.

Complete Peer review History: <http://www.sdiarticle4.com/review-history/61039>

Original Research Article

Received 07 July 2020
Accepted 13 September 2020
Published 25 September 2020

ABSTRACT

This study examines the sensitivity of the posterior mean to change in the prior assumptions. Three plausible choices of prior which include informative, relative-non informative and non-informative priors are considered. The paper considers information level for a prior to cause a notable change in the Bayesian posterior point estimate. The study develops a framework for evaluating a bound for a robust posterior point estimate. The Ellipsoid Bound theorem is employed to derive the Ellipsoid Bound for an independent normal gamma prior distribution. The proposed modification ellipsoid bound for the large prior was established by varying different variance co-variance sizes for the independent normal gamma prior. This bound represents the range for the posterior mean when is insensitive and when it's sensitive in both location and spread. The result shows that; for a large prior parameter value (greater than the OLS estimate) with a positive definite prior variance covariance matrix, and prior parameter values interval which contains the OLS estimate then, the posterior estimate will be less than both the OLS and the prior estimates. Similarly, if the lower bound of the prior parameter values range is greater than the OLS estimate then: The posterior

*Corresponding author: E-mail: muhdbashola@yahoo.com;

estimate will be greater than the OLS estimate but smaller than the prior estimate. Furthermore, it is observed that no matter the degrees of confidence in the prior values, data information is powerful enough to modify it.

Keywords: Bayesian posterior estimate; precision parameter; robustness; independent normal gamma distribution.

1. INTRODUCTION

To resolve that the use of prior information is restricted to Bayesian methods which is to consider researchers preconceived idea and position of believe, before the observation or experiment, about population of which he seeks knowledge [1,2]. The prior even guide how the experiment is carried out or observed, to create the experimental evidence (likelihood) about the uncertainty of interest θ [3,4].

The conclusion is given as a result of support for or against the researchers believes (prior) in the presences of observed data.¹ In the application of the Bayesian method of analysis [5,6,7] about an unknown parameter, it is required that quantification of the prior information should be in form of prior distribution. Importance of the prior, relevance of a posterior estimate in the presence of the OLS estimate and to some extent, the sensitivity of results to change in the prior assumptions (precision, sample size, prior degrees of freedom, distribution, functional form, etc) are always recognized in Bayesian paradigm [8,9]. These concerns are of a great interest with certain validity, as it is required for a sound Bayesian report, thus the need of proper elicitation of these conditions. In practice, availability of prior information, no matter how relevant appears improper, due to the limitation of proper and accurate quantification in terms of distribution and functional form for the Prior information.

No matter how modern operational definitions of personal probability may look, it is usually possible to determine the personal probabilities of important events only crudely [10]. Hence it is unreasonable to expect that beliefs can be model by a single Prior distribution. But, the prior distribution in itself represent a degree of uncertainty in the data, hence its uncertainty in

itself can then be altered by the information contained in the data [11]. As long as prior information is not too vague (Large variance), it is altered only by the signals and not by the noise contained in the sample [12]. Thus after deciding a single prior $p(\beta)$, through a carefully carried out process of prior elicitation, one would usually feel somewhat uncertain about the posterior $p(\beta/y)$ as any other prior, that is close to $p(\beta)$ will seems to be equally plausible.

In this study, some prior convictions were highlighted; issues regarding the sensitivity of posterior results to the prior assumptions [13] are also discussed. However, robust priors [14] are required for an Ellipsoid bound theorem to function properly. Basically, this theorem needs a positive definite prior variance co-variance to explain the relationship of the prior, posterior and OLS estimates [15,10,16].

The rest of the paper is divided into four sections. Section 2 discusses the theoretical framework for the study. Section 3 shows the result of the simulations and the data analysis. Section 4 summarises the modification of the Ellipsoid bound theory. Concluding remarks are given in section 5.

2. THE THEORETICAL FRAMEWORK FOR THE STUDY

2.1 The Model

The linear regression model is the workhorse of econometrics. The linear regression model, presents a linear relationship between the dependent variable and a $1 \times k$ vector of explanatory variables x_i , where y are indexes of the relevant observational unit. In matrix notation, the linear regression model can be written as

$$y = X\beta + \varepsilon \tag{1}$$

where: $\beta = [\beta_0, \beta_1, \beta_2, \dots, \beta_k]$, is a $(k+1)$ vector of regression parameters, X , is an $(n \times (k+1))$ Matrix of Explanatory variables Y , $1 \times n$ vector of response variable (data) $\varepsilon \sim N_m(0_n, \sigma^2 I)$ is an N vector of errors which are independently

¹ Coherence requirements lead one to believe that, given a sampling model, the only sensible way to make inference about its parameters is to assess a prior distribution describing ones initial knowledge about their values and use the data to derive, via Bayes theorem, the appropriate posterior distribution [17]

Multivariate Normally distributed, With mean 0_n , an n vector of zeros, I_n is an $n \times n$ identity matrix. $h = \sigma^{-2}$, is the error precision and thus, the normal linear regression model depends on the parameter vector $[\beta, h]$.

Assumptions about ε and X which define the likelihood function:

- Independency of ε and X .
- X are exogenous.
- Elements of X are not correlated.

2.2 The Likelihood

By using the properties of the multivariate normal distribution, it follows that $p(y|\beta, h) = \varphi(y; X\beta, h^{-1}I_n)$, and thus the likelihood function is given by;

$$P(y/\beta, h) = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} \exp \left[-\frac{h}{2} (y - X\beta)'(y - X\beta) \right] \quad (2)$$

2.3 Bayesian Inference with the Exponential Family Sampling Model

Consider the exponential family likelihood (e.g Normal distribution), for a random variable Y whose probability distribution depends on a single parameter $\theta = X\beta$.

The distribution can be written in the form

$$p(y_i/\theta) = h(y_i) \exp[\theta y_i - g(\theta)], \quad \text{for } i = 1, 2, \dots, n. \quad (3)$$

Summing over n , the functions h and g are known and y_n denotes the full vector of n observations.

It is straightforward to show that the MLE is determined by solving

$$g_0(\hat{\theta}_n) = \sum_{i=1}^n y_i(n^{-1}) = \bar{y}_n \quad (4)$$

Where $\hat{\theta}_n$, represent the MLE. That is $g_0(\theta_0)$, from eqn 4, can be interpreted as the population

$$\sqrt{n}[E(\psi/y) - \psi_0] = \sqrt{n}[w_n \hat{\psi}_n + (1 - w_n) \underline{\mu}_\psi - (w_n + (1 - w_n) \psi_0)] \quad (10)$$

Since

$$w_n + (1 - w_n) = 1$$

Hence:

$$\sqrt{n}[E(\psi/y) - \psi_0] = w_n \sqrt{n}(\hat{\psi}_n - \psi_0) + \sqrt{n}(1 - w_n)(\underline{\mu}_\psi - \psi_0) \quad (11)$$

average of the random variable y and our maximum likelihood estimate is derived from the sample analog of this moment condition.

The Bayesian takes the exponential family likelihood above, add to it a prior (in this case, a Natural conjugate prior), and combines them via Bayes theorem to obtain a posterior distribution for θ .

The prior of such family is given by:

$$p(\theta/\underline{a}, \underline{b}) \propto \exp[\underline{a}\theta - \underline{b}g(\theta)] \quad (5)$$

\underline{a} and \underline{b} are hyper parameters selected by the researcher. Using the same type of argument that was used to derive eqn 4, it is straightforward to show, provided the moment exists,

$$E(\psi) = \underline{\mu}_\psi = \underline{a}\underline{b}^{-1} \quad (6)$$

With $\psi \equiv g_0(\theta)$, Combining the likelihood derived from eqn 3 where $\sum_{i=1}^n y_i = n\bar{y}_n$ with the prior in eqn 5, we obtain via Bayes theorem

$$p(\theta/y_n) \propto \exp[(n\bar{y}_n + \underline{a})\theta - (n + \underline{b})g(\theta)] \quad (7)$$

Inspection of Equation 7 reveals that the prior in eqn 5 is indeed conjugate, as the posterior in eqn 7 is of the same functional form as the prior, with updated definitions of the parameters $\bar{a} = n\bar{y}_n + \underline{a}$ and $\bar{b} = n + \underline{b}$. Using the result in Equation 6, we then obtain

$$E(\psi/y_n) = (\bar{a}/\bar{b}) \quad (8)$$

From the above equation let $w_n = \bar{a}/\bar{b}$ then eqn 8 is equivalently as:

$$E(\psi/y_n) = w_n \hat{\psi}_n + (1 - w_n) \underline{\mu}_\psi \quad (9)$$

Which gives a weighted average of the data mean (or MLE $\hat{\psi}_n = \bar{y}_n$) and the prior mean $\underline{\mu}_\psi$ with the weight w_n . eqn 9 immediately reveals

Obviously, as $w_n \rightarrow 1$ the Bayesian posterior mean and the frequentist MLE in this case are asymptotically equivalent. That is, the asymptotic sampling properties of the Bayesian posterior mean are identical to the sampling properties of the classical MLE. The above result position the Frequentist who questions the role of the prior in a potentially insecure position, as it is clear that the weight of information needed in eqn 11 can be shared between the data and the prior information. To conduct an inference, for which finite-sample results are seldom available, the Frequentist typically relies on asymptotic approximations to the sampling distribution of the estimator. Also, under this large-sample metric, the sampling distribution of the Bayes rule is identical, suggesting that, according to his recipe for an inference, the prior does not matter.

Hence as shown above Bayesian posterior intervals with large samples generally will enjoy good Frequentist coverage probabilities, whereas, numerically, the reported classical confidence interval should be close to the Bayesian posterior interval. Never the less, since realization of large samples is seldom available for a true life situation, but how large the samples, is unknown.

2.4 Assessing the Effect of Prior Variance Assumption on the Sensitivity of Posterior Means

The model used in simulating the data is considered as the base line model, denoted as M_0 , characterized by the Likelihood $p(y/\theta, M_0)$ and the prior as $p(\theta/M_0)$ which yield a posterior distribution $p(\theta/y, M_0)$, where θ includes $\bar{\beta}$ (vector of β s) and h (precision parameter). Of interest to this research is to determine how sensitive, the reported estimate of posteriors mean (a typical Bayesian point estimate) is to changes in the prior assumptions (prior values). A possible way to achieve it is to separately re-estimate the models under the different prior Assumptions, obtain simulations from this new posterior distribution, and use these simulations to recalculate the posterior mean. This procedure is however, unappealing, as one may desire to consider a wide range of possible Variation in Prior Assumption, because considerable effort and computing time will be required to fit the model over and over again.

Another related procedure in the likeness of importance sampling [18], is simply to reweight

the simulations from the initial baseline model to assess the impact of the prior change. We explain below why such an approach is valid, and how it is often implemented in practice.

Let consider a different model, denoted as M_1 , that contains the same parameters, y and likelihood function that characterize the baseline model M_0 , but with a different prior $p(\theta/M_1)$. Then the posterior mean under this new prior is obtained as.

$$E(\theta/y, M_1) = \frac{\int_{\theta} \theta p(y/\theta, M_1) p(\theta/M_1) d\theta}{\int_{\theta} p(y/\theta, M_1) p(\theta/M_1) d\theta} \tag{12}$$

where the denominator is the marginal distribution of the data, it serves as the normalizing constant of the posterior distribution. Suppose the baseline posterior $p(\theta/y, M_0)$, were to be used as an importance function to numerically approximate values of both the numerator and denominator integrals in Equation 12. The insight of importance sampling is to divide and multiply terms within the integrand by another density from which draws are easily obtained, to enable the application of direct Monte Carlo integration. The choice of the baseline posterior as the importance function affords some considerable simplifications to this general exercise, as the likelihoods are unchanged in M_1 and M_0 , and the normalizing constant of the baseline posterior cancels in the ratio of Equation 12. Considering this, the desired posterior mean can be written as

$$E(\theta/y, M_1) = \frac{\int_{\theta} \theta \{p(\theta/M_1)/p(\theta/M_0)\} p(\theta/y, M_0) d\theta}{\int_{\theta} \{p(\theta/M_1)/p(\theta/M_0)\} p(\theta/y, M_0) d\theta} \tag{13}$$

Since:

$p(y/\theta, M_1) = p(y/\theta, M_0) = p(\theta/y, M_0)/p(\theta/M_0)$. The advantage of Equation (13) is that the averaging within the integrals is done now with respect to the baseline posterior $p(\theta/y, M_0)$ for which a set of simulations is already available. As a result, one does not need to re-estimate the model to assess prior sensitivity, but instead can simply reweight the baseline posterior simulations in the appropriate way. Specifically, the forms of the integrals in Equation 13 suggest that a simulation-consistent estimate of the new posterior mean is

$$E(\theta/y, M_1) = \sum_{m=1}^M \theta_0^{(m)} \omega_m \tag{14}$$

Where the weight ω_m are defined as

$$\omega_m \equiv \frac{p[\theta_0^{(m)}/M_1]/p[\theta_0^{(m)}/M_0]}{\sum_{m=1}^K p[\theta_0^{(m)}/M_1]/p[\theta_0^{(m)}/M_0]} \tag{15}$$

where 'k' is the number of θ in the model.

The above is used to check posterior mean changes with changes in the prior by simply reweighting the initial $\{\theta_0^{(m)}\}_{m=1}^M$. Simulations that are more likely to arise under the new prior are appropriately assigned more weight in the calculation.

2.5 The Prior and Posterior

2.5.1 The prior

An independent Normal Gamma prior is used for this study.

$$P(\beta, h) = P(\beta) \cdot P(h), \tag{16}$$

$$p(\beta, h/y) \propto \exp \left[-\frac{1}{2} \left\{ (\beta - \bar{\beta})' \bar{V}^{-1} (\beta - \bar{\beta}) \right\} \right] \exp \left[-\frac{1}{2} Q \right] h^{\frac{N+v-2}{2}} \exp \left(\frac{-hv}{2s^{-2}} \right) \tag{18}$$

Where:

$$Q = hy'y + \beta' \underline{V}^{-1} \beta - \beta' \underline{V}^{-1} \bar{\beta}$$

The joint posterior density in equation (18) is not a, well known distribution form, but holding each of the parameter constant makes the full conditionals obtainable. But approximately:

$p(\beta, h/y) \approx p(\beta/y, h) p(h/y, \beta)$ according Law of Large Numbers, which states that; if adequate random alternating samples can be obtained from the full conditionals, then our mean estimate can approximate the true posterior density estimate. Thus, the implementation of a posterior simulator called The Gibbs sampler which simulate from each parameter given that the others have occurred will be adopted.

The respective full conditionals are:

- Ignoring terms that do not involve β (including Q), we can write

$$p(\beta|h, y) \propto \exp \left[-\frac{1}{2} (\beta - \bar{\beta})' \bar{V}^{-1} (\beta - \bar{\beta}) \right] \tag{19}$$

which is the kernel of multivariate Normal density. In other words

$$\beta|y, h \sim N(\beta, V) \tag{20}$$

$$P(\beta) = \frac{1}{(2\pi)^{\frac{k}{2}}} |V|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\beta - \underline{\beta})' V^{-1} (\beta - \underline{\beta}) \right]$$

$$P(h) = C_G^{-1} h^{\frac{v-2}{2}} \exp \left(\frac{-hv}{2s^{-2}} \right) \tag{17}$$

Where,

where, k = prior sample size, C_G^{-1} is an integrating constant, $E[\beta/y] = \underline{\beta}$, prior mean of β

$Var(\beta/h) = \underline{V}$, the prior covariance matrix of β with the mean of h as \underline{s}^{-2} and \underline{V} is the degree of freedom.

The degree of information about prior parameter β is captured in the variance co-variance matrix \underline{V} , in that the larger it is the lesser the information about the prior parameters.

2.5.2 The posterior

In this study, the posterior distribution doesn't have a familiar form, hence the posterior properties can not be obtained analytically. It is usually denoted by $P(\beta, h/y)$ mathematically, using $p(\beta, h/y) = p(y/\beta, h)p(\beta, h)$ which gives

- Ignoring terms that don't involve h, we have

$$p(h|y, \beta) \propto h^{\frac{N+v-2}{2}} \exp \left[-\frac{h}{2} \{ (y - X\beta)'(y - X\beta) + v s^2 \} \right] \quad (21)$$

which is the Kernel of a Gamma density. Therefore,

$$h|y, \beta \sim G(s^{-2}, \bar{v}) \quad (22)$$

2.6 Ellipsoid Bound Theorem, Ellipsoid Modifications, Bartlett's Paradox for a Conjugate Normal Gamma Distribution for Prior

Introducing these concepts under a different umbrella of a conjugate Normal Gamma Prior distribution, such that our parameters estimation can first be understood analytically and with simple extension [19], its validity can be checked with the simulation method of the non-analytic posterior distributions. This is necessary because if the analytical agrees then the simulated methods should also do.

2.6.1 Ellipsoid bound theorem

Ellipsoid bound Theorem states that If there exist a random variable $y \sim NG(\beta, h^{-1}V)$ where β is a vector of parameters and V is the variance co variance Matrix, with natural conjugate prior which is a Normal - Gamma Distribution: $\beta, h \sim NG(\beta, \underline{V}, \underline{s}^{-2}, \underline{v})$,

for every positive definite prior Variance co variance matrix \underline{V} then:

$$(\bar{\beta} - \beta_{ave})' X' X (\bar{\beta} - \beta_{ave}) \leq \frac{(\hat{\beta} - \underline{\beta})' X' X (\hat{\beta} - \underline{\beta})}{4} \Rightarrow (\underline{\beta} \leq \bar{\beta} \leq \hat{\beta}) \quad (23)$$

Where:

$$\beta_{ave} = \frac{1}{2} (\hat{\beta} + \underline{\beta})$$

Where, $\hat{\beta}$ = OLS estimate, $\underline{\beta}$ = prior parameter and $\bar{\beta}$ = posterior parameter.

Recall, the Posterior is given as $p(\beta, h|y) = NG(\bar{\beta}, \bar{V}, \bar{s}^{-2}, \bar{v})$

With, $V = (\underline{V}^{-1} + h(X'X))^{-1}$ and $\beta = V(\underline{V}^{-1}\beta + h(X'X)\beta)$

Equation (23) is proved as follow: if $\beta = V(\underline{V}^{-1}\beta + (X'X)\beta_b)$, then from above, it can be seen that

$$(\bar{\beta} - \underline{\beta})' \underline{V}^{-1} (\bar{\beta} - \underline{\beta}) = (\bar{\beta} - \underline{\beta})' X' X (\hat{\beta} - \bar{\beta}) \quad (24)$$

Since \underline{V}^{-1} is positive definite, then $(\bar{\beta} - \underline{\beta})' \underline{V}^{-1} (\bar{\beta} - \underline{\beta}) > 0$, hence from the equation (24) above it can be concluded that:

$$(\bar{\beta} - \underline{\beta})' X' X (\hat{\beta} - \bar{\beta}) \geq 0 \quad (25)$$

From the Ellipsoid Bound equation above expanding the bracket in the L.H.S and R.H.S of equation (25), and collecting like terms

$$\begin{aligned} (\bar{\beta} - \underline{\beta})' X' X (\bar{\beta} - \hat{\beta}) &\leq 0 \\ (\bar{\beta} - \underline{\beta})' X' X (\hat{\beta} - \bar{\beta}) &\geq 0 \end{aligned}$$

which satisfied the bound in equation (25), where $\underline{\beta}, \hat{\beta}, \bar{\beta}$ are all (k x 1) vectors and $X'X$ is a (k x k) matrices.

If $k = 1$, $X'X = \sum_{i=1}^n X^2$

Therefore,

$$(\bar{\beta} - \underline{\beta})' X' X (\hat{\beta} - \bar{\beta}) = (\bar{\beta} - \underline{\beta}) \sum_{i=1}^n X^2 (\hat{\beta} - \bar{\beta}) = (\bar{\beta} - \underline{\beta})' (\hat{\beta} - \bar{\beta}) \geq 0$$

Given the two equations:

$$\begin{aligned} (\bar{\beta} - \underline{\beta}) &\geq 0 \dots\dots\dots (i) \\ (\hat{\beta} - \bar{\beta}) &\geq 0 \dots\dots\dots (ii) \end{aligned}$$

Therefore: $\bar{\beta} \geq \underline{\beta}$ and $\hat{\beta} \geq \bar{\beta}$

which implies

$$\underline{\beta} \leq \bar{\beta} \leq \hat{\beta} \tag{26}$$

hence proved.

The above explains that for any given positive prior variance covariance, the posterior estimate are always less than the OLS estimate.

2.6.2 Theoretical modification of the ellipsoid bound for large prior

Case 1:

Large prior $\underline{\beta}^*$ with positive definite \underline{V}^* , such that the lower bound of the interval created by it for the prior parameter $\underline{\beta}^*$ is $\geq \hat{\beta}$, such that equation (25) becomes

$$(\bar{\beta} - \underline{\beta}^*)' X' X (\hat{\beta} - \bar{\beta}) \leq 0 \tag{27}$$

As result of the prior $\underline{\beta}^*$ used in the Ellipsoid bound theorem then equation (27) becomes,

$$(\bar{\beta} - \beta_{ave})' X' X (\hat{\beta} - \beta_{ave}) \geq \frac{(\bar{\beta} - \underline{\beta}^*)' X' X (\bar{\beta} - \underline{\beta}^*)}{4}, \beta_{ave} = \frac{1}{2} (\hat{\beta} + \underline{\beta}^*), \tag{28}$$

If equation (28) is expanded, it gives back equation (27),

Similarly, when $k = 1$: $X'X = \sum_{i=1}^n X^2$

Therefore,

$$(\bar{\beta} - \underline{\beta}^*)' X' X (\hat{\beta} - \bar{\beta}) = (\bar{\beta} - \underline{\beta}^*) \sum_{i=1}^n X^2 (\hat{\beta} - \bar{\beta}) = (\bar{\beta} - \underline{\beta}^*)' (\hat{\beta} - \bar{\beta}) \leq 0 \tag{29}$$

Gives the two equations:

$$(\bar{\beta} - \underline{\beta}^*) \leq 0 \tag{i}$$

$$(\hat{\beta} - \bar{\beta}) \leq 0 \tag{ii}$$

Therefore: $\bar{\beta} \leq \underline{\beta}^*$ and $\hat{\beta} \leq \bar{\beta}$, which implies

$$\hat{\beta} \leq \bar{\beta} \leq \underline{\beta}^* \tag{30}$$

Case 2:

Large prior $\underline{\beta}^*$ with positive definite \underline{V}^* , such that the lower bound of the interval created by it for the prior parameter $\underline{\beta}^*$ is $\leq \hat{\beta}$, such that equation (25) becomes

$$(\bar{\beta}^{**} - \underline{\beta}^*)'X'X(\hat{\beta} - \bar{\beta}) \leq 0 \tag{31}$$

Since $(\bar{\beta}^{**} - \underline{\beta}^*) \leq 0$. As a result of the prior $\underline{\beta}^*$, and variance covariance matrix \underline{V}^* used, the Ellipsoid bound becomes;

$$(\bar{\beta}^{**} - \beta_{ave})'X'X(\bar{\beta}^{**} - \beta_{ave}) \geq \frac{(\hat{\beta} - \underline{\beta}^*)'X'X(\hat{\beta} - \underline{\beta}^*)}{4} \tag{32}$$

By expansion, it becomes,

$$(\bar{\beta}^{**} - \underline{\beta}^*)'X'X(\hat{\beta} - \bar{\beta}^{**}) \leq 0 \tag{33}$$

Thus, equation (33) satisfied equation (31) above, then similarly,

If k = 1: $X'X = \sum_{i=1}^n X^2$

Therefore,

$$\begin{aligned} (\bar{\beta}^{**} - \underline{\beta}^*)'X'X(\hat{\beta} - \bar{\beta}^{**}) &= (\bar{\beta}^{**} - \underline{\beta}^*) \sum_{i=1}^n X^2 (\hat{\beta} - \bar{\beta}^{**}) \\ &= (\bar{\beta}^{**} - \underline{\beta}^*)' (\hat{\beta} - \bar{\beta}^{**}) \leq 0 \end{aligned} \tag{34}$$

Thus, equation (34) gives these two equations:

$$(\bar{\beta}^{**} - \underline{\beta}^*) \leq 0 \tag{i}$$

$$(\hat{\beta} - \bar{\beta}^{**}) \geq 0 \tag{ii}$$

Therefore: $\bar{\beta}^{**} \leq \underline{\beta}^*$ and $\hat{\beta} \geq \bar{\beta}^{**}$, which implies

$$\bar{\beta}^{**} \leq \hat{\beta} \leq \underline{\beta}^* \tag{35}$$

2.6.3 Bartlett's paradox

Given two models: M_0 = Unconstrained model and M_1 = Constrained model

The posterior odd ratio given as

$$K_{01} = \frac{P(M_0/y_i)}{P(M_1/y_i)} \tag{36}$$

Though posterior estimates may not be affected by the Prior variance covariances, but excessive large size of it will always yield support for the unconstrained model. Also of a suitable benefit to an inference, is to be able to select the best model with minimum parameters possible, as it is cleared that all independent variables don't have equal effects on the dependent variable [20].

3. ANALYSIS OF RESULTS

To illustrate the effect of prior assumptions for the posterior sensitivity, the degree of confidence (DC) in the prior information about h, which is defined as the ratio of the prior degrees of freedom (\underline{v}) to the sample size (n), $DC = \frac{\underline{v}}{n}$. Under each cases, two possible situations of Priors are considered; (i) prior greater than the OLS Estimate and (ii) Prior Less than OLS Estimate and for both cases change in variance assumption was considered. Data simulation study was carried out to show how close the estimates are to the true parameters.

$$y_i = \sum_{j=0}^3 \beta_j X_j + \varepsilon_i \tag{37}$$

where $\varepsilon_i \sim N(0,1)$, $j = 0,1,2,3$, $i = 1, \dots, n$, $\beta_s = 1, 10, 28, 100$, $X_0 = 1$, $X \sim \text{unif}(0,1)$ As a result of the form the posterior distribution is taken, it seems impossible to obtain posterior estimate for the parameters by direct simulation, hence, the Gibbs sampler method is introduced, to simulate from full conditionals for the estimate. As a result of the posterior simulator employed, the

parameter of interest β is a function of the Data and precision parameter, from the full conditionals $p(\beta|y, h)$, where the precision depends on the posterior degrees of freedom ($\underline{v} + N$) and precision mean (s^{-2}) therefore, the interest is to observe the effect of change in parameter assumptions (degree of freedom of h and variance covariance matrix).

3.1 Accessing Sample Size Effect on Ellipsoid Bound Theorem

Case 1: In this case, the effect of change in prior and weighted prior Variance covariance matrix using relatively small data information is checked. $DC = 0.6$; the prior information about 'h' has about 60% weight, as the data information.

Table 1 shows that, for small sample sizes, the prior position with respect to the OLS estimate have no effect on posterior position, as the posterior estimates with respect to the two prior, though not equal but remains \leq the OLS result, while increase in prior variance (10^{100}) yields equal posteriors which is also \leq the OLS estimate, showing a conformity to the ellipsoid bound and also with this increase, have little or no effect on the Highest Posterior Density Interval (HPDI).

Case 2: Increasing the sample sizes, such that $DC = 0.15$; 15% prior information about h is about of the weight as data information. The Precision prior mean (\underline{s}^{-2}) was also varied. Table 2, shows that, variation in the prior mean of the precision parameter (h), only affects the HPDI and have little or no effect on the posterior estimate.

Table 1. Sample size = 5, $\underline{v} = 3$ & $DC = 0.6$

		Case 1			
		β_0	β_1	β_2	β_3
	$\hat{\beta}$ (SE)	-0.180(0.08)	8.94(0.074)	32.40(0.09)	97.86(0.07)
	Tr β	1	10	28	100
	$\underline{\beta}_1$	0	5	15	50
	$\underline{\beta}_2$	5	15	50	120
$I_{v1} = 10$	$\bar{\beta}_{11}$ (SE)	-0.095(0.005)	8.88(0.005)	32.36(0.006)	97.76(0.004)
	HPDI(.95)	(-0.83, 0.60)	(8.14, 9.55)	(31.49, 33.25)	(97.08, 98.41)
	$\bar{\beta}_{21}$ (SE)	-0.20(0.005)	8.96(0.005)	32.40(0.006)	97.89(0.004)
	HPDI(.95)	(-0.91, 0.51)	(8.29, 9.70)	(31.52, 33.29)	(97.25, 98.57)
$I_{v2} = 10^{100}$	$\bar{\beta}_{12}$ (SE)	-0.18(0.005)	8.94(0.005)	32.40(0.006)	97.86(0.004)
	HPDI(.95)	(-0.89, 0.53)	(8.26, 9.68)	(31.54, 33.26)	(97.19, 98.50)
	$\bar{\beta}_{22}$ (SE)	-0.18(0.005)	8.94(0.005)	32.40(0.006)	97.87(0.005)
	HPDI(.95)	(-0.89, 0.53)	(8.25, 9.67)	(31.53, 33.28)	(97.23, 98.56)

Table 2. Sample size = 20, $\nu = 3$ & DC= 0.15

		Case 2			
		β_0	β_1	β_2	β_3
	$\hat{\beta}$ (SE)	-0.180(0.08)	8.94(0.074)	32.40(0.09)	97.86(0.07)
	Tr β	1	10	28	100
	β_1	0	5	15	50
	β_2	5	15	50	120
$I_{\nu_1} = 10$	$\bar{\beta}_{11}$ (SE)	1.35(0.006)	8.57(0.007)	28.54(0.006)	99.47(0.006)
	HPDI(.95)	(0.48, 2.22)	(7.45, 9.64)	(27.48, 29.55)	(98.57, 100.40)
	$\bar{\beta}_{21}$ (SE)	1.16(0.006)	8.64(0.007)	28.62(0.007)	99.73(0.006)
$I_{\nu_2} = 10^{100}$	HPDI(.95)	(0.27, 2.01)	(7.59, 9.73)	(27.6, 29.62)	(98.81, 100.62)
	$\bar{\beta}_{12}$ (SE)	1.23(0.04)	8.62(0.05)	28.55(0.05)	99.71(0.04)
	HPDI(.95)	(-4.74, 7.40)	(1.18, 16.08)	(20.90, 35.14)	(92.97, 105.55)
	$\bar{\beta}_{22}$ (SE)	1.36(0.24)	8.42(0.22)	28.72(0.22)	99.45(0.27)
	HPDI(.95)	(-8.24, 12.29)	(-3.51, 20.96)	(16.08, 40.08)	(88.65, 108.51)

Case 3: Increasing the sample size = 50, keeping degrees of freedom $\nu = 3$, such that $DC = 0.06$; 6% prior information about h is about of the weight as data information.

Table 3 shows that, increasing the sample size, with small variance covariance, makes a prior greater than the OLS estimate, yield a posterior estimates approximately equal to the OLS estimate and it reduces the HPDI hence increasing the precision.

Case 4: Increasing the sample size = 100 and holding the prior degrees of freedom constant at 3, such that $DC = 0.03$; 3% prior information about precision parameter h.

Table 4, shows that, increasing the sample size, with small variance covariance, makes a prior greater than the OLS estimate, yield a posterior estimate approximately equal to the OLS estimate and it reduces the HPDI hence increasing the precision.

3.2 Accessing Variance Covariance effect on Large Prior

Since the posterior mean is the weighted average of the prior values and the OLS estimate, posteriors are expected to take values between them. However, not every posterior mean based on informative prior lies between the OLS estimates and the Prior values.

Table 3. Sample size = 50, $\nu = 3$ & DC= 0.06

		Case 3			
		β_0	β_1	β_2	β_3
	$\hat{\beta}$ (SE)	1.26(0.43)	8.84(0.52)	28.85(0.48)	99.78(0.46)
	Tr β	1	10	28	100
	β_1	0	5	15	50
	β_2	5	15	50	120
$I_{\nu_1} = 10$	$\bar{\beta}_{11}$ (SE)	1.53(0.008)	8.79(0.010)	28.80(0.010)	99.36(0.010)
	HPDI(.95)	(0.29, 2.77)	(7.31, 10.39)	(27.38, 30.18)	(98.92, 100.67)
	$\bar{\beta}_{21}$ (SE)	1.15(0.008)	8.90(0.010)	28.9(0.010)	99.89(0.009)
$I_{\nu_2} = 10^{100}$	HPDI(.95)	(0.29, 2.97)	(7.41, 10.29)	(27.38, 30.18)	(98.52, 100.97)
	$\bar{\beta}_{12}$ (SE)	1.26(0.008)	8.84(0.010)	28.88(0.010)	99.78(0.009)
	HPDI(.95)	(0.02, 2.49)	(7.30, 10.33)	(27.46, 30.21)	(98.43, 101.04)
	$\bar{\beta}_{22}$ (SE)	1.25(0.008)	8.83(0.009)	28.87(0.009)	99.80(0.009)
	HPDI(.95)	(-0.02, 2.46)	(7.21, 10.28)	(27.46, 30.24)	(98.52, 101.22)

Table 4. Sample size = 100, $\nu = 3$ & DC= 0.03

		Case 4			
		β_0	β_1	β_2	β_3
$l_{v1} = 10$	$\hat{\beta}$ (SE)	1.22(0.28)	9.29(0.31)	28.44(0.31)	99.88(0.33)
	$Tr\beta$	1	10	28	100
	$\underline{\beta}_1$	0	5	15	50
	$\underline{\beta}_2$	5	15	50	120
	$\overline{\beta}_{11}$ (SE)	1.76(0.01)	9.28(0.01)	28.31(0.01)	98.95(0.01)
$l_{v2} = 10^{100}$	HPDI(.95)	(-0.03, 3.40)	(7.48, 11.12)	(26.29, 30.02)	(96.84, 96.84)
	$\overline{\beta}_{21}$ (SE)	1.03(0.01)	9.33(0.01)	28.66(0.01)	100.07(0.01)
	HPDI(.95)	(-0.63, 2.74)	(7.59, 11.28)	(26.77, 30.55)	(98.08, 102.09)
	$\overline{\beta}_{12}$ (SE)	1.20(0.01)	9.30(0.01)	28.43(0.01)	99.92(0.01)
	HPDI(.95)	(-0.63, 2.95)	(7.45, 11.34)	(26.48, 30.35)	(97.83, 101.92)
	$\overline{\beta}_{22}$ (SE)	1.20(0.01)	9.29(0.01)	28.46(0.01)	99.90(0.01)
	HPDI(.95)	(-0.63, 2.94)	(7.42, 11.29)	(26.47, 30.38)	(97.86, 101.90)

If X is normally distributed with mean μ and variance σ^2 , then $Z = \frac{x-\mu}{\sigma}$ is standard normal, with $E(Z) = 0$ and $var(Z) = 1$. Since the prior information about the parameters is, $\beta \sim N(\beta, \underline{V}_{vi})$, therefore:

$$(\beta - \underline{\beta})(\underline{V}_{vi})^{-1} \sim N(0, 1) \tag{38}$$

A 95% confidence interval for the parameter β is given as:

$$\underline{\beta} \pm 1.96 \sqrt{(\underline{V}_{vi})} \tag{39}$$

Special Case 1: Keeping all parameter in case 1 constant, a large prior was introduced with four levels of variance covariance matrix, the largeness at each level, was captured in the scalar value l_{vi} .

Table 5 shows that, when the prior considered are far greater than the OLS estimates, then the interval created by the prior variance covariance matrix for the prior values, is of significant relevance to the position of the posterior estimate, for normal gamma prior with normal likelihood. When the interval does not contain the OLS estimates the posterior estimates are greater than the OLS, but less than it when the OLS estimates exists in the interval.

Special Case 2: Keeping all parameter in case1 constant, a large prior was introduced with four levels of variance covariance matrix, the largeness at each level, was captured in the scalar value l_{vi} .

Table 6, shows that, when prior considered are far greater than the OLS estimates, with large

data information, the interval created by the prior variance covariance matrix for the prior values, remains significant, with respect to the position of the posterior estimate, for normal gamma prior with normal likelihood. When the interval lower bound is greater than the OLS estimates the posterior estimates are greater than the OLS (small prior variance), $\hat{\beta} \leq \underline{\beta} \leq \beta$, but less than it when the OLS estimates is greater than that its lower bound of the interval (Large prior variance covariance), $\overline{\beta} \leq \hat{\beta} \leq \beta$.

Considering all the cases shown above, it is observed that the degree of confidence in Prior information have no effect on the posterior estimate and when the OLS estimate exist in the interval created by the prior variance covariance matrix or the prior is less than the OLS estimate, then posterior estimate will be less than the OLS Estimate and with greater precision.

3.3 Prior Effect on Negligible Data

Special Case 3: In this special case, it is to explore how sensitive the posterior estimate can be to negligible data information. This may be seen as an impossible situation but it can suggest a solution to a limitation. Here we consider data of sample size = 3 with prior degrees of freedom taken as 30. Where the confidence in the prior takes a positive value of DC=10. In the table below it was observed that no matter how small or negligible data might appear, combination of such to a prior can yield a reasonable posterior estimate. The HPDI contains the true parameter of interest when the prior is reasonably far from the true parameter.

Attaching a prior with large variance covariance, to negligible data information leads to loss of the whole information.

confidence in the prior takes a positive value of 5; DC = 5.

Special Case 4: This examines the sensitivity of the posterior estimate, considering small data information with large prior degrees of freedom. Here data of sample size = 5 were selected, with prior degrees of freedom taken as 25 and

Table 8 depicts that no matter how relevant the prior information might be for a normal Likelihood with an independent normal Gamma Prior, the data information is reasonably powerful enough to enable the prior yields a reasonable posterior estimate.

Table 5. Sample size = 5, $\nu = 3$ & DC= 0.6

		Special case 1				
		β_0	β_1	β_2	β_3	
$l_{v1} = 10$	$\hat{\beta}$ (SE)	-0.1803 (0.0750)	8.9403 (0.0745)	32.3984 (0.0914)	97.8612 (0.0703)	
	Tr β	1	10	28	100	
	β_1	50	100	280	1000	
	$\bar{\beta}_{11}$ (SE)	-1.0979 (0.0048)	9.8378 (0.0047)	32.2192 (0.0058)	99.7515 (0.0048)	
	HPDI(.95)	(-1.825, -0.368)	(9.139, 10.571)	(31.346, 33.101)	(99.048, 100.493)	
	$l_{v2} = 100$	$\bar{\beta}_{12}$ (SE)	-0.2784(0.0048)	9.0346(0.0047)	32.3867(0.0058)	98.0511(0.0044)
	$l_{v3} = 1000$	$\bar{\beta}_{13}$ (SE)	-0.1968(0.0046)	8.9544(0.0046)	32.3982(0.0057)	97.8901(0.0044)
	$l_{v4} = 10^{100}$	$\bar{\beta}_{14}$ (SE)	-0.1862(0.0047)	8.94674(0.0047)	32.4017(0.0058)	97.8621(0.0044)
		HPDI (.95)	(-0.888, 0.557)	(8.211, 9.632)	(31.529, 33.300)	(97.205, 98.529)

Table 6. Sample size = 100, $\nu = 3$ & DC= 0.03

		Special case 2				
		β_0	β_1	β_2	β_3	
$l_{v1} = 10$	$\hat{\beta}$ (SE)	1.2214 (0.2840)	9.2896 (0.3083)	28.4468 (0.3075)	99.8843(0.3293)	
	Tr β	1	10	28	100	
	β_1	50	100	280	1000	
	$\bar{\beta}_{11}$ (SE)	-4.1663 (0.0252)	9.3760 (0.0098)	29.3804 (0.0104)	109.98 (0.0471)	
	HPDI(.95)	0.3173(0.0119)	9.3070(0.0124)	28.6351(0.0125)	101.52(0.0147)	
	$l_{v2} = 100$	$\bar{\beta}_{12}$ (SE)	1.1118(0.0116)	9.2944(0.0127)	28.4861(0.0127)	100.05(0.0137)
	$l_{v3} = 1000$	$\bar{\beta}_{13}$ (SE)	1.2018(0.0118)	9.3013(0.0127)	28.4279(0.0127)	99.9157(0.0137)
	$l_{v4} = 10^{100}$	$\bar{\beta}_{14}$ (SE)	(-0.6289, 2.9481)	(7.4452, 11.3446)	(26.4887, 30.357)	(97.83, 101.924)

Table 7. Sample size = 3, $\nu = 30$ & DC= 10

		Special case 3				
		β_0	β_1	β_2	β_3	
$l_{v1} = 10$	$\hat{\beta}$ (SE)	120.46(NA)	-110.32(NA)	32.41(NA)	NA(NA)	
	Tr β	1	10	28	100	
	β_1	0	5	15	50	
	β_2	5	15	50	120	
	$\bar{\beta}_{11}$ (SE)	5.85(0.03)	2.97(0.03)	32.50(0.01)	92.81(0.03)	
	HPDI(.95)	(1.25, 11.03)	(-1.83, 7.82)	(31.17, 33.77)	(88.85, 96.88)	
	$\bar{\beta}_{21}$ (SE)	1.03(0.01)	9.33(0.01)	28.66(0.01)	100.07(0.01)	
	HPDI(.95)	(-0.63, 2.74)	(7.59, 11.28)	(26.77, 30.55)	(98.08, 102.09)	
	$l_{v2} = 10^{100}$	$\bar{\beta}_{12}$ (SE)	NA(NA)	NA(NA)	NA(NA)	NA(NA)
		HPDI(.95)	NA(NA)	NA(NA)	NA(NA)	NA(NA)
	$\bar{\beta}_{22}$ (SE)	NA(NA)	NA(NA)	NA(NA)	NA(NA)	
	HPDI(.95)	NA(NA)	NA(NA)	NA(NA)	NA(NA)	

Note: NA = Not Available

Table 8. Sample size = 5, $\nu = 25$ & DC= 5

		Special case 4			
		β_0	β_1	β_2	β_3
$I_{\nu_1} = 10$	$\hat{\beta}$ (SE)	-0.180(0.08)	8.94(0.074)	32.40(0.09)	97.86(0.07)
	Tr β	1	10	28	100
	$\underline{\beta}_1$	0	5	15	50
	$\underline{\beta}_2$	5	15	50	120
	$\overline{\beta}_{11}$ (SE)	-0.04(0.006)	8.84(0.006)	32.34(0.008)	97.69(0.006)
$I_{\nu_2} = 10^{100}$	HPDI(.95)	(-0.96, 0.86)	(7.86, 9.70)	(31.22, 33.50)	(96.84, 98.56)
	$\overline{\beta}_{21}$ (SE)	-0.21(0.006)	8.97(0.006)	32.41(0.007)	97.92(0.006)
	HPDI(.95)	(-1.13, 0.71)	(8.16, 9.92)	(31.29, 33.54)	(97.08, 98.80)
	$\overline{\beta}_{12}$ (SE)	-0.18(0.006)	8.95(0.006)	32.40(0.007)	97.86(0.006)
	HPDI(.95)	(-1.07, 0.77)	(-1.07, 0.77)	(8.05, 9.89)	(97.03, 98.73)
	$\overline{\beta}_{22}$ (SE)	-0.18(0.006)	8.94(0.006)	32.41(0.008)	97.86(0.006)
	HPDI(.95)	(-1.10, 0.76)	(8.05, 9.89)	(31.30, 33.55)	(97.03, 98.76)

4. MODIFICATION OF THE ELLIPSOID BOUND FOR THE INDEPENDENT NORMAL GAMMA PRIOR

- For a large prior value which is greater than the OLS estimate, if there exist a positive definite prior variance covariance matrix \underline{V} , with an interval for the prior values, which contains the OLS Estimate, Then:

$$\overline{\beta} \leq \hat{\beta} \leq \underline{\beta}$$

- For a large prior value which is greater than the OLS estimate, if there exist a positive definite prior variance covariance matrix \underline{V} , with an interval for the prior values, which the lower bound is greater than the OLS Estimate, then:

$$\hat{\beta} \leq \overline{\beta} \leq \underline{\beta}$$

5. CONCLUSION

This study has shown the asymptotic equivalence of the Frequentist MLE and the Bayesian Posterior mean. The result shows that, for a large prior parameter value (greater than the OLS estimate) with a positive definite prior variance covariance matrix and prior parameter values interval which contains the OLS estimate then, the posterior estimate will be less than both the OLS and the prior estimates. Similarly, if the lower bound of the prior parameter value range is greater than the OLS estimate then: the posterior estimate will be greater than the OLS estimate but smaller than the prior estimate. Finally, it was

shown that relevant data information under Multivariate Normal distribution serves as a good modifying factor for the independent Normal Gamma prior to yield reasonable posterior estimates; mean closer to the true value, smaller posterior variance realized. Also, the higher the prior variance covariance is, the better estimates are obtained as the sample size increases especially when data are negligible. The Credible intervals (HPDI) for the posterior estimates give smaller/closer posterior intervals due to the smaller error variances when compared with the confidence intervals of OLS estimates which can only be varried by data information.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Girshick MA, Mosteller F, Savage LJ. Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statist.* 1946;17:13-23.
2. Hedibert F. Lopes, Justin L. Tobias. Confronting prior convictions: On issues of prior sensitivity and likelihood robustness in Bayesian analysis. *The Annual Review of Economics Journal*; 2011.
3. Berger JO. *Statistical decision theory and Bayesian analysis*. Second Edition. New York: Springer-Verlag; 1987.
4. Berger JO. The case for objective Bayesian analysis. *Bayesian Anal.* 2006;1:385402.
5. Bernardo J, Smith AFM. *Bayesian theory*. Chichester: John Wiley and Sons; 1994.

6. Bernardo JM, Ramn JM. An introduction to Bayesian reference analysis. *The Statistician*. 1998;47:101135.
7. Gary K. Bayesian econometrics. England. Wiley; 2003.
8. Escobar M, West M. Bayesian density estimation and inference using mixtures. *Journal of American Statistics Association*. 1995;90:57788. Prior and Likelihood Sensitivity. 129. Available:www.annualreviews.org
9. Lopes HF, West M. Bayesian model assessment in factor analysis. *Stat. Sinica*. 2004;14:4167.
10. Savage, Leonard J. The foundations of statistics reconsidered. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Contributions to the Theory of Statistics*, University of California Press, Berkeley, Calif. 1961;1:575-586.
11. Jim A. Bayesian computation with R. Second Edition, Springer, Bowling Green State University, Bowling OH, USA. 2009;117-147.
12. Mattaio C, Alessandro R. Bayesian VARs: A survey of the recent literature with an application to the European Monetary System IMF Working Paper; 2003.
13. Olubusoye OE, Okewole DM. Prior sensitivity in Bayesian linear regression model. *STARS: Int. Journal (Sciences)*. 2009;3(1):21-29.
14. Berger JO. Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference*. 1990;25:303328.
15. Nurit Olikier, Avi Ostfeld. Minimum volume ellipsoid classification model for contamination event detection in water distribution systems. *Procedia Engineering*. 2014;70:1280–1288.
16. Lamperski J, Freund RM, Todd MJ. An oblivious ellipsoid algorithm for solving a system of (ln) feasible linear inequalities. arXiv preprint arXiv 1910 03114, arxiv.org; 2019.
17. Jose MB. Reference posterior distributions for Bayesian inference. *J. R. Statis. Soc. B*. 1979;419(2):113-147.
18. Geweke J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*. 1989;57:13171340.
19. Yushuang Liu, Yan Zhao, Falin Wu. Extended ellipsoid outer-bounding set-membership estimation for nonlinear discrete-time system with unknown-but-bounded disturbances. *Discrete Dynamics in Nature and Society*; 2016.
20. Bartlett MS. Comment on D. V. Lindleys statistical paradox. *Biometrika*. 1957;44: 53334.

APPENDIX

1. DATA SIMULATION

```

1 #for B_1,V*10 Bprior 1
2 setwd("C:/users/DELL/Desktop/R Training")
3 #-----Data Simulation-----
4 set.seed(5)
5 x1=runif(3,0,1)
6 set.seed(5)
7 set.seed(6)
8 x2=runif(3,0,1)
9 set.seed(6)
10 set.seed(7)
11 x3=runif(3,0,1)
12 set.seed(7)
13 X=cbind(1,x1,x2,x3)
14 head(X)
15 set.seed(5)
16 E=rnorm(3,0,1)
17 Bs=as.matrix(rbind(1,10,28,100))
18 Y=X%%Bs+E
19 head(Y)
20 summary(lm(Y~X))

```

2. PRIOR SPECIFICATION AND INITIALIZATION

```

21 #-----prior B variance, expected value prior precision,
22 #-----and prior precision degree of freedom
23 N=5
24 v_ =3
25 spi_2 = .0025
26 Bpr1=as.matrix(rbind(0,5,15,50))
27 V_p=matrix(c(1,0,0,0,0,2,0,0,0,0,5,0,0,0,0,5),nrow=4,
28           ncol=4,byrow=T)
29 IVI1=10*diag(4)#weighing matrix
30 V_pri1=v_p%%IVI1
31 #-----Initialization-----
32 mcmc = 10000
33 samples <- matrix(0,mcmc)
34 sampled.B11=samples%%rep(1,4)
35 sampled.B110 = numeric(mcmc)
36 sampled.B111 = numeric(mcmc)
37 sampled.B112 = numeric(mcmc)
38 sampled.B113 = numeric(mcmc)
39 sampled.h11 = numeric(mcmc)
40 current.B11 =as.matrix(cbind(0,10,25,50))#initial B for simulatic
41 current.h11 = 3 #initial h for simulation
42 sampled.B11[1,]=current.B11#vector of beta
43 sampled.B110[1] = sampled.B11[1,1]
44 sampled.B111[1] = sampled.B11[1,2]

```

3. GIBBS SAMPLER

```

45 sampled.B112[1] = sampled.B11[1,3]
46 sampled.B113[1] = sampled.B11[1,4]
47 sampled.h11[1] = current.h11
48 #-----Gibbs sampler-----
49 for(i in 2:mcmc){
50   vpst11=solve(solve(v_pri1)+drop(sampled.h11[i-1])*t(X)%*%X)
51   Bpost11 = vpst11%*(solve(v_pri1)%*(Bpri1) +
52     drop(sampled.h11[i-1])*t(X)%*%Y)
53   # sample from full conditionals of B
54   library(MASS)
55   current.B11= mvrnorm(1, Bpost11, vpst11)
56   K11=as.matrix(current.B11)
57   # sample from the full conditional of phi
58   V_pst=v_+N
59   SSE11=(t(Y-X%*%K11)%*(Y-X%*%K11))
60   s_pst11=(SSE11+(v_*spi_2^1))/V_pst
61   current.h11 = rgamma(1, s_pst11, V_pst) #s_pst11= variance of t
62 #-----RESULTS 11-----
63 sampled.B110[i] = K11[1,]
64 sampled.B111[i] = K11[2,]
65 sampled.B112[i] = K11[3,]
66 sampled.B113[i] = K11[4,]
67 sampled.h11[i] = current.h11
68 }

69 #-----Burn In-----
70 bin=4000
71 fin.B110 = sampled.B110[-(1:bin)]
72 fin.B111 = sampled.B111[-(1:bin)]
73 fin.B112 = sampled.B112[-(1:bin)]
74 fin.B113 = sampled.B113[-(1:bin)]
75 fin.h11 = sampled.h11[-(1:bin)]
76 #==== Posterior ESTIMATIONS 12=====
77 mean(fin.B110)
78 sd110=sd(fin.B110)
79 SE110=sd110/sqrt(6000)
80 SE110
81 mean(fin.B111)
82 sd111=sd(fin.B111)
83 SE111=sd111/sqrt(6000)
84 SE111
85 mean(fin.B112)
86 sd112=sd(fin.B112)
87 SE112=sd112/sqrt(6000)
88 SE112
89 mean(fin.B113)
90 sd113=sd(fin.B113)
91 SE113=sd113/sqrt(6000)
92 SE113]

```

© 2020 Akanbi et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
 The peer review history for this paper can be accessed here:
<http://www.sdiarticle4.com/review-history/61039>