Ilorin Journal of Science
Volume 1, Number 1, 2014, pp. 216 – 227 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2014 Faculty of Physical Sciences, university of Ilorin
https://doi.org/10.54908/iljs.2014.01.01.019

ILORIN
JOURNAL OF SCEINCE

ILJS-14-014

On Bayesian Conjugate Normal Linear Regression and Ordinary Least Square Regression Methods: A Monte Carlo Study

Yahya[1], W.B., Olaniran[1], O.R. and Ige[2], S.O.

[1]Department of Statistics, University of Ilorin, PMB 1515, Ilorin, Nigeria
[2]Department of Mathematics and Statistics, Federal Polytechnic, Ado-Ekiti, Nigeria

Abstract

In this study, comparison between the classical ordinary least square (OLS) regression technique and the Bayesian conjugate normal linear regression method when the data satisfy all the necessary assumptions of OLS technique is presented. The Bayesian normal linear regression model was fitted using Normal-Gamma conjugate prior. Results from Monte Carlo study showed that the OLS estimator is as good as the Bayesian estimator in terms of the closeness of their estimated parameters to the true values. However, using the criteria of the mean square errors of parameters' estimates and other performance indices, the results showed that Bayesian estimator is more efficient, more consistent and relatively more stable than the classical least squares method even when the sample data satisfy all the necessary assumptions of the OLS method. The apparent better performance of Bayesian estimator over the OLS is justified by the prior information about the data that Bayesian technique employed in its estimation. Therefore, it could be concluded that if reliable information about the data under investigation is available, Bayesian regression technique(s) that would make use of such information should be preferred for efficient model's estimation and better inference. The R statistical package (www.cran.org) was employed for all the analysis in this study.

Keywords: Bayesian conjugate normal linear regression, ordinary least squares, Normal-Gamma conjugate prior, Mean Square Error, Credible interval

1. Introduction

A number of real life phenomena often call for the establishment of some form of relationships among several measurable quantities to aid meaningful judgement. Such a relationship is simply formalized through the concept of (linear and non-linear) regression techniques (Yahya and Olaifa, 2014). Consider a multiple linear regression model that connects a quantitative response of interest y and a set of metrical and/or categorical covariates $x_0, x_1, \ldots, x_k$ of the form

$$y = x_0\beta_0 + x_1\beta_1 + \ .\ .\ . + x_k\beta_k \ + \varepsilon \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ is a vector of regression parameters to be estimated and ε is the random error term of the model with the assumption that $\varepsilon \sim N(0, \sigma^2)$. In order to have an

Corresponding Author: Yahya, W.B.
Email: wb_yahya@daad-alumni.de

intercept model, the value of x_0 in equation (1) is simply set to 1 across all the samples. Within the frequentist framework, a celebrated method of the ordinary least squares (OLS) (Gujarati, 2004) is usually an appealing technique to estimate the regression model in (1) when the underlying assumptions for its usage are met by the data. Some of these assumptions are that of the independent samples, constancy of error variance and normality of model error term among others (Greene, 2000; Yahya et al., 2008). The OLS technique simply minimizes the error sum of squares of the regression model and has been reported to be the *best linear unbiased estimator* (BLUE) of the regression model in (1) (Gujarati, 2004).

Under Bayesian reasoning, whenever relevant prior information and belief about the behaviour of the situation being modelled are available and known to the investigator, it may be desirable to make use of such information in the estimation of the regression model in (1). Making use of such information is expected to improve the efficiency of the estimated regression model (Barber, 2012).

If the sample data satisfy all the necessary conditions that underlie the use of the least square technique, the OLS estimator becomes more efficient and would produce the best linear unbiased estimates of the regression parameters. However, if prior information about, say, the functional forms of the model's parameters exists for example, the use of such information in model's estimation may influence the behaviour of the estimators.

The present work therefore, is intended to examine the performance of both the frequentist (OLS) and Bayesian estimators for modelling linear regression model (1) given that all the assumptions that underlying the use of the OLS estimator are met and that prior information on the functional forms of the regression parameters is available. While the Bayesian estimator would make use of the known prior information to improve its performance, the OLS technique is expected to efficiently model the data since all the necessary conditions required for its usage are met by the data. Using a Monte Carlo approach, the assessment criteria of the mean square error, the 95% confidence and credible intervals of parameters' estimates shall be employed to evaluate the performances of the two forms of estimation at varying sample sizes.

2. Materials and Methods

In this section, brief overview of the classical OLS and Bayesian conjugate normal linear regression techniques for estimating the regression model (1) are provided.

Let the matrix representation of the linear regression model (1) be given by

$$y = X\beta + \varepsilon \tag{2}$$

where y is the $n \times 1$ vector of responses, X is the $n \times p$ matrix of predictor variables, β is the $p \times 1$vector of the regression coefficients while ε is the $n \times 1$ vector of random noise of the model that is assumed to have independent Gaussian density with zero mean and a constant variance σ^2, i.e. $\varepsilon \sim N(0, \sigma^2)$. The normality assumption on ε gave the name *normal linear regression model* often call model (2) (Gujarati, 2004).

2.1 *Brief Overview of the OLS Estimation*

Traditionally, the least square estimator of the linear regression model (2) seeks to minimize the residual sum of square $SSE = \sum_{i=1}^{n}(y_i - X_i\hat{\beta})^2$ in model (2). The estimated vector $\hat{\beta}$ that minimizes parameter β is called the least squares estimate of β and is simply computed by

$$\hat{\beta}_{ols} = (X^T X)^{-1}(X^T y) \tag{3}$$

The 100(1- α)% confidence interval on $\hat{\beta}_{ols}$ at Type I error rate α is obtained as

$$\hat{\beta}_{ols} \pm (t_{1-\frac{\alpha}{2},\ n-k} \times SE(\hat{\beta}_{ols})) \tag{4}$$

where $SE(\hat{\beta}_{ols}) = \sqrt{\widehat{\sigma^2}(X^T X)^{-1}}$ and $\widehat{\sigma^2}$ is computed by

$$\widehat{\sigma^2} = \frac{1}{n-k} SSE = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - X_i\hat{\beta})^2 \tag{5}$$

2.2 *Bayesian Conjugate Linear Regression Model Estimation*

Estimation of the linear regression model (2) using Bayesian techniques can be performed through the following three steps (Simon, 2009);

i.) Determine the likelihood function of the unknown parameters to be estimated given the data.

ii.) Specify the prior for all the unknown parameters.

iii.) Determine the posterior distribution of the parameters given the data.

2.2.1 *The Likelihood Function*

Suppose $Y_1, \dots, Y_n$ be a set of i.i.d random samples of size n from discrete/continuous density function $f(y_i; \theta)$ with an unknown parameter θ. Then, the likelihood function of θ is given by $L(\theta) = \prod_{i=1}^{n} f(y_i; \theta)$.

It is noted, from the linear regression model $y = X\beta + \varepsilon$ in (2) that $\varepsilon \sim N(0, \sigma^2 I_n)$. This shows that $y \sim N(X\hat{\beta}, \sigma^2 I_n)$ and if we set $var(\varepsilon) = \sigma^2 I_n = h^{-1} I_n$, we have the representation that

$$y \sim N(X\hat{\beta}, h^{-1} I_n) \tag{6}$$

where I_n is the unit matrix of dimension n. Thus, $\varepsilon \sim N(0, h^{-1} I_n)$ and $h = \sigma^{-2}$. The notation that $var(\varepsilon) = h^{-1} I_n$ is a compact notation for $var(\varepsilon_i) = h^{-1}$ and $cov(\varepsilon_i, \varepsilon_j) = 0$ for $i, j = 1, \dots . n$ and $i \neq j$.

All elements in predictor matrix X of model (2) are either fixed (i.e not random variables) or random. If they are random variables, they are independent of all elements of ε with a probability density function $p(X|\gamma)$ where γ is a vector of parameters that does not include β and $h(\text{Lindley and Smith, 1972})$.

The above statement implies that we can proceed by not conditioning on X and treat $L(\beta, h|y)$ as the likelihood function of parameter vector $\omega = (\beta, h)$ from the joint density function $p(y|X, \beta, h)$. We drop X from the conditioning set to simplify the notation. From the representation in (6) and using the definition of multivariate normal density, we can write the likelihood function of y as:

$$L(\beta, h|y) = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} \left\{ exp\left[\frac{-h}{2} (y - X\beta)^T (y - X\beta) \right] \right\} \tag{7}$$

The product $(y - X\beta)^T (y - X\beta)$ in (7) can be expressed in terms of the OLS estimator b of β as:

$$(y - X\beta)^T (y - X\beta) = (y - X\beta + Xb - Xb)^T (y - X\beta + Xb - Xb) \tag{8}$$

where $b = (X^T X)^{-1} (X^T y)$. Thus, we have

$$(y - X\beta)^T (y - X\beta) = (y - Xb)^T (y - Xb) + (b - \beta)^T (X^T X)(b - \beta) \tag{9}$$

Recall that the estimator s^2 of the variance of the model σ^2 is

$$s^2 = \frac{(y - Xb)^T (y - Xb)}{n - k} \tag{10}$$

Thus,

$$(n - k)s^2 = (y - Xb)^T (y - Xb) \tag{11}$$

Substituting (9) and (11) in (7) to have

$$L(\beta,h|y) \; = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}}\left\{exp\left[\frac{-h}{2}\left((n-k)s^2+(b-\beta)^T(X^TX)(b-\beta)\right)\right]\right\}$$

and setting $v = n - k$, which is interpreted as the degree of freedom in the above to have

$$L(\beta,h|y) \; = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}}\left\{exp\left[\frac{-h}{2}(vs^2+(b-\beta)^T(X^TX)(b-\beta))\right]\right\}. \qquad (12)$$

If (12) is partitioned by using $n = v + k$, the likelihood function for $\omega = (\beta, h)$ becomes

$$L(\beta,h|y) \; = \frac{1}{(2\pi)^{\frac{n}{2}}}\left\{h^{\frac{k}{2}}\,exp\left[\frac{-h}{2}(b-\beta)^T(X^TX)(b-\beta)\right]\right\}\left\{h^{\frac{v}{2}}\,exp\left[\frac{-hv}{2s^{-2}}\right]\right\}. \qquad (13)$$

The quantity $\left\{h^{\frac{k}{2}}\,exp\left[\frac{-h}{2}(b-\beta)^T(X^TX)(b-\beta)\right]\right\}$ in (13) resembles the kernel of the multivariate Gaussian density while $\left\{h^{\frac{v}{2}}\,exp\left[\frac{-hv}{2s^{-2}}\right]\right\}$ also looks like the kernel of the gamma density. These results simply suggest a *normal-gamma* prior for the likelihood function (Koop, 2003).

2.2.2 *Specifications for Priors and Their Distributions*

Priors are meant to reflect any information the researcher has before seeing the data which he wishes to incorporate in the analysis of the data. Hence, priors can take any form. However, it is common to choose particular classes of priors that are easy to interpret and/or which would make computation easier (Gelman, 2006). *Natural conjugate priors* typically have both such advantages.

A *conjugate prior distribution* is one which, when combined with the likelihood, yields a posterior that falls in the same class of distributions (Raifa and Schlaifer, 1961). A natural conjugate prior has the additional property that it has the same functional form as the likelihood function. This property means that the prior information can be interpreted in the same way as the likelihood function's information. In other words, the prior can be interpreted as arising from a fictitious dataset from the same process that generated the actual data.

For the linear regression model (2), we must elicit a prior distribution for β and h, which we denote by $p(\beta,h)$. The fact that we are not conditioning on the data means that $p(\beta,h)$ is a prior density, the posterior density will be denoted by $p(\beta,h|y)$. It proves convenient to write $p(\beta,h|y) = p(\beta|h)p(h)$ and think in terms of a prior for $p(\beta|h)$ and one for h. The form of the likelihood function in (13) suggests that the natural conjugate prior will involve a Gaussian distribution for $\beta|h$ and a Gamma distribution for h. This is indeed the case. The name given to a distribution such like this which is a product of a Gamma and a (conditional) Normal is the *Normal-Gamma density*.

Therefore, it follows that

$$\beta|h\sim N(\beta_0, h^{-1}\sum\nolimits_0)$$

and

$$h\sim Gamma\left(s_0{}^{-2}, \frac{2s_0{}^{-2}}{v_0}\right).$$

Hence, the distribution of the natural conjugate prior of both β and h is given as:

$$\beta, h\sim NG(\beta_0, \sum\nolimits_0, s_0{}^{-2}, v_0),$$

a normal–gamma density. It then follows that

$$p(\beta|h) = \frac{1}{2\pi^{\frac{k}{2}}|h^{-k}\sum_0|^{\frac{1}{2}}}\left\{exp\left[\frac{-1}{2}(\beta-\beta_0)^T(h^{-1}\sum\nolimits_0)^{-1}(\beta-\beta_0)\right]\right\}. \tag{14}$$

Note that, $|h^{-k}\sum_0| = h^{-k}|\sum_0|$ and that $(h^{-1}\sum_0)^{-1} = h(\sum_0)^{-1}$ where $|.|$ denotes the determinant of the argument.

Hence, (14) becomes

$$p(\beta|h) = \frac{h^{\frac{k}{2}}}{2\pi^{\frac{k}{2}}|\sum_0|^{\frac{1}{2}}}\left\{exp\left[\frac{-h}{2}(\beta-\beta_0)^T(\sum\nolimits_0)^{-1}(\beta-\beta_0)\right]\right\}, \tag{15}$$

also,

$$p(h) = \frac{1}{\Gamma\left(\frac{v_0}{2}\right)\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}h^{\frac{v_0}{2}-1}\,exp\left(\frac{-hv_0}{2s_0{}^{-2}}\right). \tag{16}$$

In the density functions (15) and (16), β_0 represents the prior mean for β, $\sum_0$ is the un-scaled variance-covariance matrix for β, while $s_0{}^{-2}$ and v_0 are the prior mean and prior degree of freedom of the gamma distribution for the model precision *h* respectively.

From Bayes theorem, it is noted that

$$p(\beta, h) = p(\beta|h)p(h) \tag{17}$$

Therefore, from (15) and (16) the density function of the conjugate prior for β and h can be written as

$$p(\beta,h)=\frac{h^{\frac{k}{2}}}{2\pi^{\frac{k}{2}}|\Sigma_0|^{\frac{1}{2}}}\left\{exp\left[\frac{-h}{2}(\beta-\beta_0)^T(\Sigma_0)^{-1}(\beta-\beta_0)\right]\right\}\times\frac{1}{\Gamma\left(\frac{v_0}{2}\right)\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}h^{\frac{v_0}{2}-1}\,exp\left\{\frac{-hv_0}{2s_0{}^{-2}}\right\} \quad (18)$$

$$\rightarrow \qquad p(\beta,h)=\frac{h^{\frac{v_0+k}{2}-1}}{2\pi^{\frac{k}{2}}|\Sigma_0|^{\frac{1}{2}}\Gamma_{\frac{v_0}{2}}\left(\frac{2s_0{}^{-2}}{v_0}\right)^{\frac{v_0}{2}}}\left\{exp\left[\frac{-h}{2}(\beta-\beta_0)^T(\Sigma_0)^{-1}(\beta-\beta_0)+\frac{v_0}{s_0{}^{-2}}\right]\right\} \quad (19)$$

The above results simply imply that the distribution of the prior $p(\beta,h)$ for β and h is multivariate normal-gamma as shown in (20)

2.2.3 *The Posterior Distributions*

The posterior density summarizes all the information from the data and the prior about the unknown parameters β and h. It is proportional to the likelihood times the prior density. For the linear regression model (2), it can be easily shown that the posterior density is also of Normal-Gamma form, thus, confirming that the prior obtained in the previous section is indeed a natural conjugate prior of the parameters β and h (Koop, 2003).

The posterior density can be simply obtained from the relation below;

$$p(\beta,h|y)=\frac{p(\beta,h)p(y|\beta,h)}{p(y|X)} \quad (20)$$

$$\rightarrow \qquad p(\beta,h|y)\propto\ p(\beta,h)\times p(y|\beta,h) \quad (21)$$

The representation (12) simplifies posterior computation since one doesn’t need the marginal likelihood $p(y|X)$ before getting the posterior densities for the regression parameters.

Formally, we have a posterior of the form (Joyce, 2009)

$$\beta,h|y\sim\ NG(\beta^*,\Sigma^*,v^*,s^{-2^*})$$

where;

$\Sigma^*=(\Sigma_0{}^{-1}+X^TX)^{-1}$,

$v^*=v_0+n$, and

$$s^{-2^*}=\frac{v^*}{v_0s_0{}^2+vs+(b-\beta_0)^T(\Sigma_0{}^{-1}+(X^TX)^{-1})^{-1}(b-\beta_0)}$$

The Bayes estimator of parameter β of linear regression model (2) is therefore given by

$$\hat{\beta}_{Bayes} = \Sigma^{*}({\Sigma_0}^{-1}\beta_0 + X^T X\hat{\beta}_{ols}) \tag{22}$$

The variance and sum of squared of the residual of the model are respectively given by (Joyce, 2009);

$$\widehat{\sigma^2}_{Bayes} = \frac{v_0 {s_0}^2 + vs + (\hat{\beta}_{ols} - \beta_0)^T ({\Sigma_0}^{-1} + (X^T X)^{-1})^{-1} (\hat{\beta}_{ols} - \beta_0)}{v^*} \tag{23}$$

$$SSE_{Bayes} = v_0 {s_0}^2 + vs + (\hat{\beta}_{ols} - \beta_0)^T ({\Sigma_0}^{-1} + (X^T X)^{-1})^{-1} (\hat{\beta}_{ols} - \beta_0) \tag{24}$$

The 100(1- α)% credible interval of the Bayesian estimator (an analogous of the confidence interval under the frequentist settings) is computed by

$$\hat{\beta}_{Bayes} \pm (t_{1-\frac{\alpha}{2}, v^*} \times SE(\hat{\beta}_{Bayes})) \tag{25}$$

where; $SE(\hat{\beta}_{Bayes}) = \sqrt{var(\hat{\beta}_{Bayes})}$ while $var(\hat{\beta}_{Bayes}) = \frac{SSE_{Bayes}}{v^*-2}\Sigma^*$

2.3 *Assessment Criteria*

Two major criteria are employed to assess the performances of the OLS and Bayesian estimators of parameters of linear regression model (2). Firstly, the confidence and credible intervals of the OLS and Bayesian estimators are used to assess their efficiency. Secondly, the *mean square error* (MSE) of parameter estimates as provided by the two estimators are equally employed to assess the closeness of their estimates to the true parameter values.

The MSE is simply computed by $MSE = \frac{1}{k}\sum_i^k (\hat{\beta}_i - \beta_i)^2$, where k is the number of parameters being estimated in the model.

3.0 Simulation Study

Different data sets were generated based on the linear regression model (2) with three predictor variables. The schemes used for data simulation were similar to those employed by Yahya et al., (2008) and later adapted by Yahya and Olaifa (2014).

The three predictor variables x_1, x_2 and x_3 in the model were generated from multivariate Gaussian density with different means and specified variance-covariance matrix that maintained the absence of inter-dependency (collinearity) among the three predictors. Specifically, the correlation structure among the predictors as imposed for simulation is of the form

$$\rho = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The values of the model's parameters were set at $\beta_0 = 85, \beta_1 = 100, \beta_2 = 95$ and $\beta_3 = 215$. Thus, the response variable y in the model was simulated based on the relationship

$$y = 85 + 100x_1 + 95x_2 + 215x_3 + \varepsilon \qquad (26)$$

where ε was simulated from a unit Gaussian density ($\varepsilon \sim N(0,1)$). All the data sets were simulated at varying sample sizes n that range between 10 and 200. To ensure stability of results, all parameter estimates were performed using 1000 replicates for both the OLS and Bayesian techniques.

3.1 *Priors Specifications*

In this study, we are under the conjugate Bayesian framework in which informative prior is conjectured. Hence, the set of values specified for the parameter vector β as used for simulation shall be the prior for the parameters of the model in (26). Therefore, $\beta_{prior} = \beta$ where $\beta' = [\beta_0 = 25, \beta_1 = 15, \beta_2 = 10, \beta_3 = 35]$.

The priors $\sum_{\boldsymbol{prior}}$, s_{prior}^2 and v_{prior} for the un-scaled variance-covariance matrix for β, the prior variance of model's residual and the prior degree of freedom respectively shall be elicited by setting arbitrary values for them in the Monte Carlo study. Therefore, the following settings were used for the different priors throughout in this study;

$$\sum\nolimits_{prior1} = \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.02 & 0 & 0 \\ 0 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0.04 \end{bmatrix}$$, *the un-scaled prior variance-covariance matrix of* β.

$s_{prior}^2 = 1$,

$v_{prior} = 4$, the prior degree of freedom which is also the prior sample size, was set to be 4, since we are estimating four parameters.

The scaled prior covariance matrix of β follows from the properties of Normal-Gamma distribution which implies that the prior covariance matrix for β_{prior} have the form:

$$var(\beta_{prior}) = \left(\frac{v_{prior} * s_{prior}^2}{v_{prior} - 2}\right) * \sum\nolimits_{prior1}$$

Thus the scaled covariance matrix of β is

$$\sum_{prior2=}\left(\frac{4*1}{4-2}\right)*\begin{bmatrix}0.01 & 0 & 0 & 0\\ 0 & 0.02 & 0 & 0\\ 0 & 0 & 0.03 & 0\\ 0 & 0 & 0 & 0.04\end{bmatrix}, \textit{the scaled prior covariance matrix of } \beta.$$

Note that all the prior covariances were set to zero. This is commonly done, since it is often hard to make reasonable guesses about what they might be. It implies that prior information about what plausible values for β_j might be are uncorrelated with those for β_i for $\neq j$. In many cases, this is a reasonable assumption.

4.0 Analysis and Results

In this section, we present the results of the OLS and Bayesian estimators for the linear regression model (26) based on the Monte Carlo study.

A total of twenty different data sets were simulated at different sample sizes (n) from $n = 10$ to $n = 200$ according to the simulation scheme presented in Section 3. Both the OLS and Bayesian conjugate normal linear regression models were fitted on each data set and parameters estimates yielded by each method are reported as shown in Table 1. The MSE of the estimated parameters by both the OLS and Bayesian estimators for each data set are equally reported in Table 1. The true values of the regression parameters that were used for data simulation are also reported in parentheses in that table under each estimation technique.

Results from Table 1 showed that both the frequentist (OLS) and Bayesian estimators provided good estimates of the parameters of linear regression model (26). In all cases, the average deviation from the true parameter values, as measured by the MSE, is not up to a unit, except in two cases (at low sample sizes 10 and 20) where OLS estimator reported MSE of about one unit. Not only this, visual examination of all the results in Table 1 revealed a good agreement between the parameter estimates yielded by both estimators and the true parameter values.

However, in terms of precision, a closer look at the MSEs reported in Table 1 showed that the Bayesian estimator is more precise than the OLS estimators at all the sample sizes considered. In all cases, the Bayesian MSEs are relatively smaller than those of the OLS. These results are clearly presented by the plots of the MSEs of both the OLS (in green dotted lines) and Bayes (in red dotted lines) estimators against the various sample sizes as shown in Fig 1. It can be easily observed from Fig 1 that Bayes estimator is relatively more efficient (for providing lower MSE values) and more stable than the OLS estimator at various sample sizes.

Table 1: Table of various OLS and Bayesian parameters estimates for the linear regression model (26) at different sample sizes. The true values of the regression parameters are given in parentheses. The MSE of the two estimators are also reported. At all the sample sizes considered, the MSE of Bayesian estimator is relatively smaller than that of the OLS estimator.

Sample Size (n)	**OLS ESTIMATES**				**BAYESIAN ESTIMATES**				**OLS MSE**	**BAYES MSE**
	$\widehat{\beta_o}$ (85)	$\widehat{\beta_1}$ (100)	$\widehat{\beta_2}$ (95)	$\widehat{\beta_3}$ (215)	$\widehat{\beta_o}$ (85)	$\widehat{\beta_1}$ (100)	$\widehat{\beta_2}$ (95)	$\widehat{\beta_3}$ (215)		
10	87.2328	99.9926	95.004	214.995	85.000	99.9975	95.0075	214.997	1.24631	1.70E-05
20	82.8681	100.038	94.9987	215.006	84.999	99.9995	94.9948	215.004	1.13623	1.22E-05
30	85.0440	99.9984	95.0032	214.999	85.000	99.9985	95.0031	214.999	0.00048	3.04E-06
40	86.5237	99.9964	94.9957	215.000	85.002	99.9995	94.9987	215.001	0.58042	1.16E-06
50	83.9370	100.061	94.9997	214.997	84.999	100.009	94.9977	214.996	0.28250	7.87E-06
60	86.1784	99.994	95.0054	214.997	85.000	99.9965	95.0075	214.999	0.34716	1.74E-05
70	85.8127	99.9976	95.0009	214.998	85.000	99.9992	95.0024	214.999	0.16514	1.81E-06
80	84.6734	100.003	94.9988	215.002	84.999	99.9997	94.9982	215.001	0.02668	1.65E-06
90	84.0099	100.022	95.0016	215.000	84.997	100.003	94.9997	214.999	0.24506	7.38E-08
100	85.6569	99.9978	94.9993	215.002	85.002	99.9991	95.0005	215.009	0.10892	4.64E-07
110	84.9857	99.9971	95.0033	215.012	85.000	99.9971	95.0033	215.012	5.6E-05	5.12E-06
120	84.5849	99.9992	95.0023	215.013	84.999	99.9984	95.0015	215.009	0.04088	1.44E-06
130	85.7567	99.997	95.0002	214.999	85.004	99.9986	95.0017	215.006	0.14137	1.34E-06
140	85.8400	99.9983	94.9967	215.006	85.004	100.000	94.9983	215.014	0.17385	1.24E-06
150	85.2281	100.006	94.9976	214.999	85.001	100.001	94.9981	215.001	0.01004	1.18E-06
160	84.9027	99.9987	94.9993	215.002	84.999	99.9985	94.9991	215.027	0.00372	2.61E-06
170	84.8304	100.007	94.9989	215.000	84.999	100.003	94.9985	215.006	0.00189	6.57E-07
180	85.1916	99.9988	94.9991	215.001	85.001	99.9992	94.9995	215.014	0.09175	7.19E-07
190	85.6334	100.006	94.9979	214.997	85.005	100.019	94.9992	214.982	0.10302	1.95E-06
200	85.0212	100.006	94.9966	215.001	85.005	100.007	94.9967	215.016	0.00116	3.47E-06

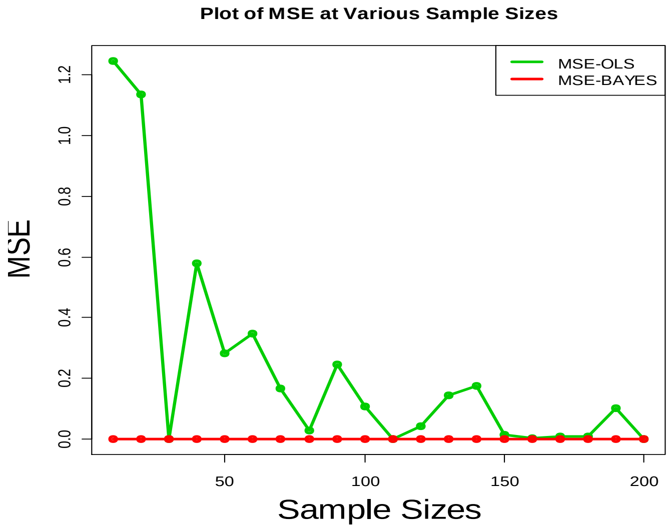
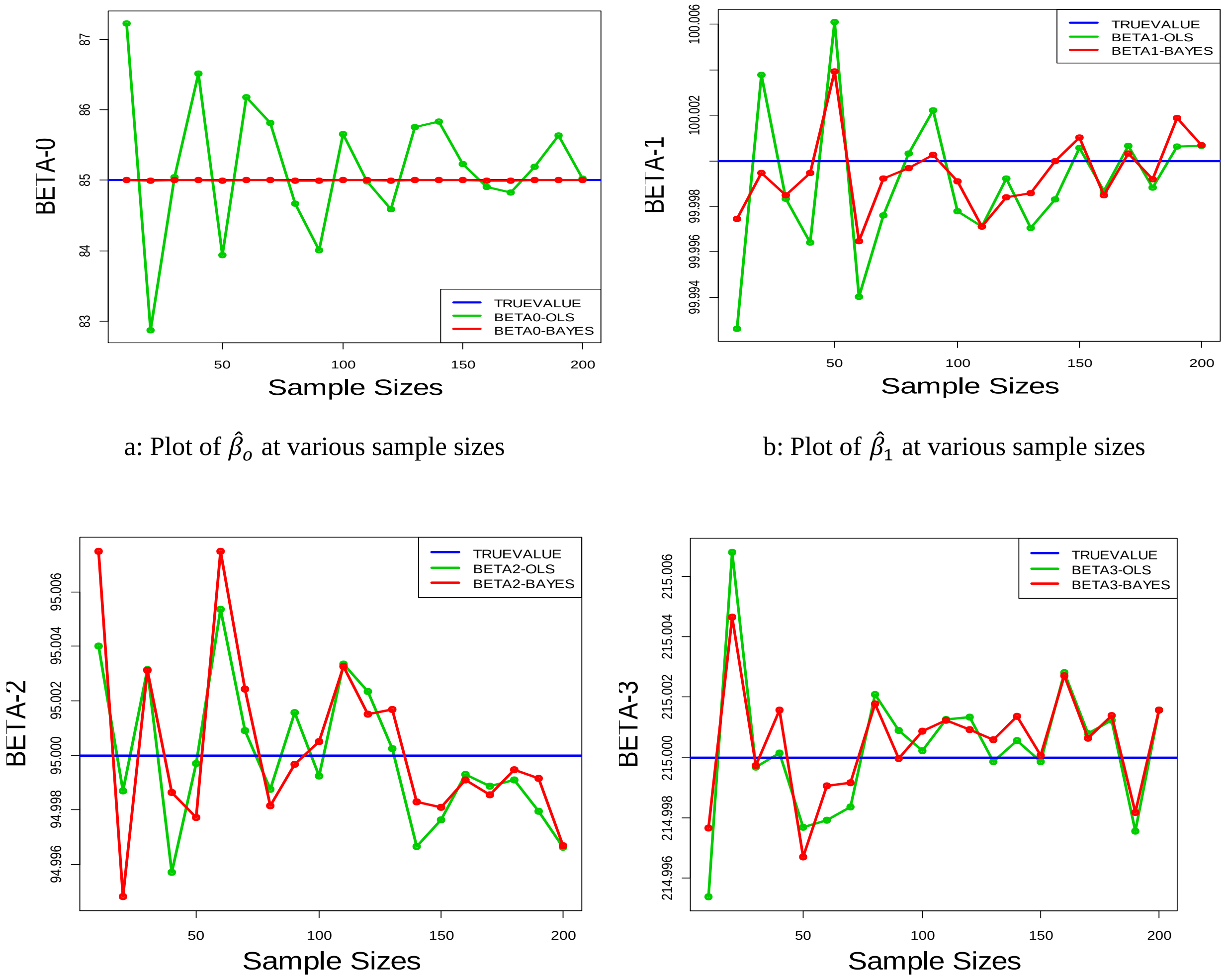


Fig 1: Line graphs of the mean square errors of regression model (26) provided by the OLS (in green dotted line) and Bayes (in red dotted line) estimators. The plots showed that the MSEs of Bayes estimator are relatively smaller and more stable than those of the OLS estimator.

To further measure the stability of the two estimation techniques, the 95% confidence and credible intervals of OLS and Bayesian estimators are respectively computed as reported in Table 2. The two results showed that Bayes estimator with relatively narrower credible interval of parameter estimates, especially on the intercept parameter, is more stable than the OLS estimators. This result is clearly evident by the plots of parameters estimates by OLS and Bayesian techniques at various sample sizes as shown as shown in Fig 2 for all the four parameters of the models β_o (written as BETA-0), β_1 (written as BETA-1), β_2 (written as BETA-2) and β_3 (written as BETA-3). The graphs showed that Bayes estimator is more stable than the OLS estimator as clearly depicted for β_o. Nevertheless, the results of the OLS estimators as shown in Table 2 are not worst off either.



a: Plot of $\hat{\beta}_o$ at various sample sizes

b: Plot of $\hat{\beta}_1$ at various sample sizes

c: Plot of $\hat{\beta}_2$ at various sample sizes

d: Plot of $\hat{\beta}_3$ at various sample sizes

Fig 2: Line graphs showing the plots of the estimated parameters (a: $\hat{\beta}_o$, b: $\hat{\beta}_1$, c: $\hat{\beta}_2$, d: $\hat{\beta}_3$) against the various sample sizes. The graphs showed that Bayes estimator that produces parameters estimates (with red lines) that are more closer to the true values (with blue lines), more noticeable for the intercept parameter $\hat{\beta}_o$, is more efficient and more stable than the OLS estimators.

Table 2: Table of the estimated 95% confidence and credible intervals by OLS and Bayesian estimators for the parameters of the linear regression model in (26). Bayesian estimator, with narrower 95% credible interval seems more stable than the OLS estimator with relatively wide 95% confidence interval especially on the intercept parameter.

Parameters	**OLS** **95% Confidence Interval**	**BAYES** **95% Credible Interval**
β_o (85)	(73.85, 98.05)	(84.69, 85.29)
β_1 (100)	(99.94, 100)	(99.95, 100)
β_2 (95)	(94.98, 95.07)	(94.99, 95.06)
β_3 (215)	(214.97, 215.02)	(214.98, 215.02)

Table 3: Table of estimated variance of the residual in the data by OLS and Bayes estimators. The results showed a reasonable agreement with the true value of the residual variance of 1 (σ^2=1).

Sample size (n)	**OLS Residual Variance**	**Bayes Residual Variance**
10	0.9867	0.7853
20	0.9589	0.8523
30	0.9533	0.8728
40	1.0171	0.9489
50	0.9975	0.9406
60	0.9708	0.9267
70	1.0042	0.9599
80	1.0120	0.9764
90	0.9681	0.9386
100	0.9997	0.9702
110	1.0243	0.9979
120	1.0104	0.9863
130	1.0122	0.9884
140	1.0074	0.9871
150	1.0083	0.9882
160	1.0124	0.9933
170	1.0121	0.9933
180	0.9934	0.9772
190	0.9897	0.9751
200	1.0131	0.9985

As a further step to examine the behaviour of the two regression estimators at retrieving the value of the residual variance, σ^2 imposed for data simulation, the OLS and Bayes residual variance at each sample size were computed. For all the simulated data, the residual variance σ^2

was fixed at one unit. That is, $\sigma^2 = 1$ for all cases. Results of the estimated variance of the residuals $\hat{\sigma}^2$ provided by the two methods as presented in Table 3 showed a good agreement of the estimated values by the two methods to the true (target) value of 1. However, a plot of these estimated variances of regression residuals against the various sample sizes revealed that Bayes estimator slightly under-estimated the true value than the OLS estimator as shown in Fig 3 while OLS slightly over-estimated the true value in a number of cases.

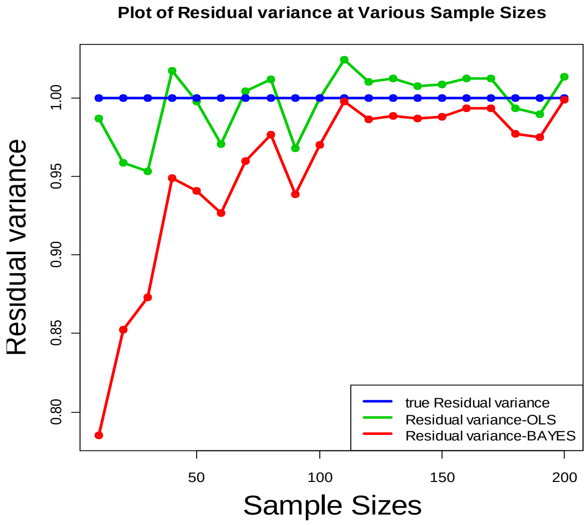


Fig 3: The plot of OLS and Bayes estimated residual variances against the sample sizes. The plots reveal that OLS technique (green dotted line) over-estimated the true variance (blued dotted line) while Bayes method (red dotted line) under-estimated the true residual variance while both of them strikes convergence as the sample size becomes larger.

Finally, the predictive strength of both the OLS and Bayes estimators of the linear regression model (26) was examined. This is achieved by comparing their mean square errors of prediction (MSEP) of the response targets y. Results showed that the two modelling techniques have similar prediction strength for the model as shown in Table 4. Table 4 presents the MSEP of the OLS and Bayes estimators for selected four sample sizes of 50, 100, 150 and 200. In all the cases, the MSEP provided by the two modelling techniques are quite close to each other.

Table 4: Table of mean square error of prediction (MSEP) provided by OLS and Bayes estimators at selected sample sizes. In all the cases, the MSEP reported by the two methods are similar.

Sample size (*n*)	**OLS MSEP**	**BAYES MSEP**
50	0.6245	0.6077
100	0.7776	0.7807
150	1.4101	1.4102
200	1.1066	1.1110

5.0 Discussions and Conclusion

The place of the Bayesian method for estimating linear regression model using informative conjugate prior vis-à-vis the classical frequentist OLS technique has been largely demonstrated in this work. Different twenty data sets were simulated through a well-structured Monte Carlo process. For easy comparison of the performances of the two modelling techniques, all the data sets were made to satisfy all the necessary conditions required for the use of the classical least square regression technique. This is intended in order to ensure that whatever results provided by the OLS technique (against which the results from Bayesian technique would be compared) is the best it could provide for such a data set.

Various results from this study showed that the two methods of estimation are quite efficient for modelling linear regression model, even when the sample data satisfy all the necessary requirements for the use of the classical least square technique. However, it is observed that the use of relevant prior information about the data in the estimation of regression model, as often the case in Bayesian methodology, would further improve the efficiency and reliability of the regression results as well as the inferences drawn from such results.

While the OLS technique was able to efficiently model the linear regression equation in (26) as expected given that the working data sets met all the necessary conditions for its usage as shown by the results in Table 1, the Bayesian conjugate normal linear regression technique was able to further improve on this level of performance (see Table 1) as a results of its ability to make use of the relevant prior information about the data.

It can be generally inferred from the various results obtained in this work that Bayesian conjugate normal linear regression technique performed creditably well like the classical frequentist technique of the OLS for modelling the regression model type in (2). In many instances, the Bayesian method outperforms its frequentist counterpart based on some of the assessment criteria. Therefore it will be of a good practice in the present of sufficient prior information, that Bayesian methodology be employed in order to guarantee a more efficient result.

Acknowledgement

We would like to appreciate the Faculty of Physical Sciences, University of Ilorin, Ilorin, Nigeria for sponsoring this maiden edition of the journal by granting waver for the Publication and Processing fees. Also, we thank the two anonymous reviewers for their suggestions and useful comments all of which have helped to improve the quality of the work.

References

Barber, D. (2012): *Bayesian Reasoning and Machine Learning*. Cambridge University Press, United Kingdom.

Gelman, A. (2006): Prior Distributions for Variance Parameters in Hierarchical Models. Bayesian Analysis, **1**(3), 515-533.

Greene, W. (2000): *Econometric Analysis,* 4th Edition: Prentice-Hall, New Jersey.

Gujarati, D. N. (2004): *Basic Econometrics*. 4th Edition, McGraw-hill, New York.

Joyce, D. (2009): A short introduction to Bayesian Statistics. *URL:* http://aleph0.clarku.edu/~djoyce/ma218/bayes1.pdf *(Accessed on January, 20, 2014).*

Koop,.G. (2003): *Bayesian Econometrics*. John Wiley & Sons Ltd., UK.

Lindley, D. V. and Smith, A. F. M. (1972): Bayes Estimates for the Linear Model. Journal of the Royal Statistical Society. Series B (Methodological), **34** (1), 1–41.

Raifa, H. and Schlaifer, R. (1961): *Applied Statistical Decision Theory*. Division of Research, Graduate School of Business Administration, Harvard University.

Simon, K. (2009), The Bayesian Linear model with unknown variance. Seminar for Statistics ETH Zurich, 9-12.

Yahya, W.B., Adebayo, S.B., Jolayemi, E.T., Oyejola, B.A. & Sanni, O.O.M. (2008): Effects of non-orthogonality on the efficiency of seemingly unrelated regression (SUR) models”. InterStat Journals, *1-29. http://interstat.statjournals.net/* ISSN: 1941-689X.

Yahya, W.B. & Olaifa, J.B. (2014): A note on Ridge Regression Modeling Techniques. *Electronic Journal of Applied Statistical Analysis*. http://siba-ese.unisalento.it/index.php/ejasa/index *(Accepted).*